

1970

# Estimators for the errors in variables model

Angel Martinez-Garza  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Statistical Theory Commons](#)

---

## Recommended Citation

Martinez-Garza, Angel, "Estimators for the errors in variables model" (1970). *Retrospective Theses and Dissertations*. 4341.  
<https://lib.dr.iastate.edu/rtd/4341>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

71-7302

MARTINEZ-GARZA, Angel, 1934-  
ESTIMATORS FOR THE ERRORS IN VARIABLES MODEL.

Iowa State University, Ph.D., 1970  
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

ESTIMATORS FOR THE ERRORS IN VARIABLES MODEL

by

Angel Martinez-Garza

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of  
The Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University  
Of Science and Technology  
Ames, Iowa

1970

## TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. REVIEW OF LITERATURE	4
III. THE ORDER IN PROBABILITY CONCEPT	7
A. Introduction	7
B. Order in Probability	11
C. Applications	15
IV. ESTIMATION OF THE SOLUTION OF LINEAR EQUATIONS WITH RANDOM COEFFICIENTS	18
A. Nearly Unbiased Estimators	18
B. Estimators with Finite Expectations	32
C. Estimation in a Simultaneous Equations Problem	41
D. Nearly Unbiased Estimators of Certain Functions of the Parameters of a Regression Model	58
V. ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR	65
A. Errors in Variables, Classical Case	65
B. The Asymptotic Bias and Variance of the Maximum Likelihood Estimators. Errors in Variables, One Variable Case	70
C. The Asymptotic Bias and Variance of the Maximum Likelihood Estimator. Errors in Variables, Several Variables Case	80
D. Comparison of the Maximum Likelihood Estimator with Other Estimators	83
VI. ESTIMATION IN A GENERALIZED COVARIANCE MODEL WITH ERRORS IN THE COVARIATES	86
A. The Covariance Model with Errors in the Covariates	86
B. The Maximum Likelihood Estimator	97

	Page
C. The Maximum Likelihood Estimator for a Simultaneous Equations Problem	99
D. The Randomized Complete Block Design with Two Covariates Subject to Error	101
VII. A SAMPLING EXPERIMENT	105
A. The Basic Theory	105
B. The Experimental Procedure	111
C. Results	117
VIII. SUMMARY	125
IX. BIBLIOGRAPHY	127
X. ACKNOWLEDGEMENTS	130

## I. INTRODUCTION

Let  $y$ , a  $n \times 1$  vector, be an observable random variable whose expectation is a known linear function of unknown parameters. That is,  $E(y) = X\beta$ , where  $X$  is a  $n \times p$  matrix of known constants whose rank is  $r \leq p$  and  $\beta$  is a  $p \times 1$  vector of unknown parameters. Assume

$$y = E(y) + e = X\beta + e, \quad (1.1)$$

where  $e$  is a  $n \times 1$  vector of random errors such that  $E(e) = 0$  and  $E(ee') = I\sigma^2$ ,  $\sigma^2$  being an unknown constant. Then it is well known that least squares gives the best linear unbiased estimator of any linear parametric "estimable" function.

Model 1.1 which is known as the General Linear Hypothesis Model, is not always the most adequate one in any situation. There are many cases where the researcher does not have control of his explanatory variables. These, instead of being measured exactly are available to the researcher containing an error of observation. In more complicated cases, the explanatory variables may be stochastic.

Consider the following situation. Let

$$\begin{aligned} Y &= y + e, \\ y &= x\beta, \\ X &= x + u, \end{aligned} \quad (1.2)$$

where  $X$  and  $x$  are  $n \times k$  matrices and  $Y$ ,  $y$  and  $e$  are  $n \times 1$  vectors. The  $x$  matrix may be either a matrix of fixed constants or a sample from a  $k$ -variate population with nonsingular covariance matrix.

Let  $(y_j, x_j)$ ,  $j = 1, \dots, n$  the  $j$ th row of the matrix  $(y, x)$ , be independent of the errors  $(e_s, u_s)$ ,  $s = 1, \dots, n$  the  $s$ th row of the matrix  $(e, u)$ , for all values of  $j$  and  $s$ . Let  $E(e_j, u_j) = (0, 0)$  for all  $j$ . Under the hypothesis that

$$\frac{1}{n} \begin{pmatrix} e'e & e'u \\ u'e & u'u \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma_e^2 & \Sigma_{eu} \\ \Sigma_{ue} & \Sigma_{uu} \end{pmatrix} \neq 0, \quad (1.3)$$

and

$$\frac{x'x}{n} \xrightarrow{P} \Sigma_{xx} \neq 0, \quad (1.4)$$

where the notation  $\xrightarrow{P}$  indicates convergence in probability and the right hand sides of 1.3 and 1.4 are positive definite matrices of corresponding orders, all of whose elements are finite constants. Then it is a simple matter to show that

$$\hat{\beta} = (X'X)^{-1}(X'Y) \xrightarrow{P} \beta + (\Sigma_{xx} + \Sigma_{uu})^{-1}(\Sigma_{ue} - \Sigma_{uu}\beta). \quad (1.5)$$

That is, the simple least squares estimator of  $\beta$  does not have the nice properties observed under the general linear hypothesis model. The simple least squares estimator is not only biased in small samples. In addition, it is an inconsistent estimator. Model 1.2 is known in the literature as the Errors in Variables Model.

One of the principal problems in connection with the Errors in Variables Model is the problem of estimation. Estimation for this kind of model depends upon the knowledge available about the covariance structure of the errors. The estimates by themselves have the mathematical structure of a generalized ratio estimator. Consequently, their small sample properties are not immediately available and the research worker

has to resort to large sample theory to study the most relevant properties of the estimators.

The most simple large sample criteria used to discriminate between proposed estimators are the asymptotic bias and variance. These two properties are defined in terms of an approximation to the true distribution of the estimators. Thus, in the absence of any other knowledge it would seem reasonable to look for statistics with the smallest asymptotic bias and variance. It was on this basis that the efforts of the present research were undertaken.

The errors in variables estimation problem is studied. The properties of several well known statistics are discussed, and compared to the alternative estimators first introduced by Fuller (1968). Additional topics relevant to the problem of estimation are also covered. In particular, a method guaranteeing the existence of the expected value of the estimators is presented. An application to the estimation in a simultaneous equations problem is treated. Also, an extension of the methodology for the covariance model with errors in the covariates, considered by DeGracie (1968), is presented. Finally, the results of a sampling experiment to study the behavior of several proposed estimators are described.



## II. REVIEW OF LITERATURE

The problem of fitting a simple linear relationship when both variables are subject to error seems to have been first studied by Adcock (1877). He defined as the line of best fit the straight line for which the sum of squares of the orthogonal distances from the sample points, is a minimum. His solution is known to be not invariant under orthogonal transformations. Kummell (1879) gave the weighted least squares solution when the variance ratio is known. His solution is identical with that given by the maximum likelihood method if the errors have the normal distribution. Kummell's estimators are consistent as well as invariant under orthogonal transformations.

After Adcock's and Kummell's first incursions into the errors in variables problem, several other papers on the subject appeared including one by Pearson (1901). However, it was not until Wald's (1940) well known paper that the subject was given serious consideration. He starts with two sets of random variables  $\{x_1, \dots, x_n\}$  and  $\{y_1, \dots, y_n\}$ ; letting  $E(x_i) = X_i$  and  $E(y_i) = Y_i$ , he defines

$$\epsilon_i = x_i - X_i, \quad i = 1, 2, \dots, n,$$

$$\eta_i = y_i - Y_i,$$

where the  $\epsilon$ 's are uncorrelated identically distributed random variables with finite variances. Similarly the  $\eta$ 's are uncorrelated identically distributed random variables with finite variances such that  $E(\epsilon_i \eta_j) = 0$ ,  $1 \leq i, j \leq n$ . Assuming that the linear relation

$$Y_i = \alpha X_i + \beta, \quad i = 1, 2, \dots, n,$$

holds between the true values, Wald's method gives consistent estimators for  $\alpha$ ,  $\beta$  and the standard errors of the  $\epsilon$ 's and  $\eta$ 's. The method requires the use of a sample of  $2m$  pairs of observations  $(x_i, y_i)$ ,  $i = 1, 2, \dots, m$ . By splitting arbitrarily the observations in two groups of size  $m$  the estimators are derived, attaining the maximum efficiency when the observations are ranked in order of magnitude. The method fails to give consistent estimators of the parameters when the errors follow the normal distribution.

A very systematic treatment of the errors in variables model was given by Lindley (1947). He discusses the conditions under which the sampling observations follow a linear regression law under the hypothesis that the true relationship is a linear function of the parameters. Once these conditions are established, the problem of estimation is attacked. The maximum likelihood solution is discussed in detail. In the one variable case, the anomalies of the maximum likelihood solution are observed. Lindley proposes that in order to obtain a more logical solution, additional information must be considered and the ratio of variances is formally introduced.

Reiersol (1945) introduced another way to arrive at consistent estimators of the parameters in question. A set of so-called "instrumental variables" highly related to the set of variables under study is used. A description of the method is given in Goldberger (1963). Later, Reiersol (1950) was able to find some connections between the errors in variables model and the identification problem. Let us briefly describe the identification subject in his own words:

When all parameters and all distributions in the model are specified, we shall talk about structure. A structure is thus a particular realization of the model, and the model is the set of all structures compatible with the given specifications. A structure generates one and only one distribution  $P(x)$  of the observed variables. If two or more structures generate the same joint probability distribution of the observed variables, the structures are said to be equivalent. If a parameter has the same value in all structures it is said to be identifiable.

Moreover,

If a parameter is not identifiable, no consistent estimate of the parameter will exist. Conversely, if an estimate of a parameter has been proved to be consistent, the parameter must be identifiable.

Some interesting remarks in connection with the subject of errors in variables are given by Kendall (1951), and (1952).

A very general method of estimation was introduced by Wolfowitz (1953). The method called "The Minimum Distance" is of wide generality of application. Wolfowitz illustrates how the method can be used to estimate the parameters of an errors in variables model.

The one variable case of the errors in variables model has been given particular attention. Madansky (1959) and Dorff (1960) give an extensive coverage of this case.

A modification to the estimator given by the method of moments has been recently proposed by Fuller (1968). This approach will be discussed in full detail in subsequent chapters. An extension of Fuller's work is one of the major topics of this dissertation.

### III. THE ORDER IN PROBABILITY CONCEPT

#### A. Introduction

We shall briefly review in this chapter the common order relationships and their connection to probability statements. Since the topic is closely related to the different types of convergence arising in probability theory, we shall consider in this introduction the most relevant concepts of the subject in relation to order relationships.

Let  $X, X_1, X_2, \dots, X_n, \dots$  be random variables defined on a probability space  $(\Omega, A, P)$ . Consider the following definitions.

Definition 1 We say that  $X_n$  converges in probability to the random variable  $X$ , and write  $X_n \xrightarrow{P} X$ , if for every  $\delta > 0$

$$\lim_{n \rightarrow \infty} P\{\omega \in \Omega ; |X_n(\omega) - X(\omega)| > \delta\} = 0 .$$

Definition 2 Given  $r > 0$ , we say that  $X_n$  converges in the  $r$ th mean to the random variable  $X$ , and write  $X_n \xrightarrow{L^r} X$ , if

$$\lim_{n \rightarrow \infty} E\{|X_n - X|^r\} = 0 .$$

Definition 3 We say that  $X_n$  converges with probability one, or almost surely to the random variable  $X$ , and write  $X_n \xrightarrow{a.s.} X$ , if

$$P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1 .$$

Finally, if  $F_1, F_2, \dots, F$  are distribution functions of some random variables over our probability space, we have:

Definition 4 We say that the sequence of distribution functions  $\{F_n\}$  converge (in distribution or in law) to a limiting probability distribution  $F$  if  $F_n$  converges to the distribution function  $F$  at every point of continuity of  $F$ . We write  $F_n \xrightarrow{d} F$ .

Let us illustrate the above concepts with two examples:

Example 1 Let  $\Omega = [0,1]$  in  $\mathbb{R}$ , with  $P$  as Lebesgue measure. Let  $F_{nj} = [(j-1)/n, j/n]$ ,  $j=1,2,\dots,n$ , and let  $\{E_n\}$  be the sequence

$$F_{11}, F_{21}, F_{22}, F_{31}, F_{32}, F_{33}, F_{41}, \dots$$

Let  $X_n = 1_{E_n}$ , the indicator function of the set  $E_n$ . Then  $P\{E_n\} \rightarrow 0$  and  $\{\omega \in \Omega : |X_n(\omega)| > 0\} = E_n$ , so that  $X_n \xrightarrow{P} 0$ . But for each  $\omega \in \Omega$ ,  $X_n(\omega) = 1$  infinitely often, so  $X_n(\omega) \rightarrow 0$  is always false. That is,  $X_n$  converges in probability to the constant random variable 0, however, it is not true that  $X_n$  converges surely to 0.

Example 2 In Example 1, let  $X_n = \frac{1}{n} \cdot 1_{E_n}$ . Again  $P\{E_n\} \rightarrow 0$  and  $\{\omega \in \Omega : |X_n(\omega)| > 0\} = E_n$ , so that  $X_n \xrightarrow{P} 0$ . Now,

$$\int_{\Omega} |X_n(\omega)| dP(\omega) \leq \frac{1}{n} \text{ for all } n,$$

which shows that  $X_n \xrightarrow{L^1} 0$ . Also for all  $\omega \in \Omega$   $X_n \rightarrow 0$  then  $X_n \xrightarrow{\text{a.s.}} 0$ .

Notice that in the above definitions we can consider convergence to the constant random variable 0. It will be sufficient to substitute 0 for  $X$ .

The result stated in the following theorem shows that convergence in mean implies convergence in probability.

Theorem 1 If  $X_n \xrightarrow{L^r} X$  then  $X_n \xrightarrow{P} X$ .

Proof It is sufficient to consider convergence to zero.

Given  $\delta > 0$ , we have

$$\begin{aligned} E\{|X_n|^r\} &= \int_{\Omega} |X_n|^r dP \\ &= \int_{\{\omega \in \Omega: |X_n| > \delta\}} |X_n|^r dP + \int_{\{\omega \in \Omega: |X_n| \leq \delta\}} |X_n|^r dP \end{aligned} \quad (3.1)$$

that the above relation is true follows from the fact that

$$\Omega = \{\omega \in \Omega: |X_n(\omega)| > \delta\} \cup \{\omega \in \Omega: |X_n(\omega)| \leq \delta\}.$$

That is,  $\Omega$  is equal to the union of the above two disjoint sets. The total integral is therefore, the sum of the partial integrals. Since both integrals on the right hand side of 3.1 are nonnegative, it follows that

$$E\{|X_n|^r\} \geq \int_{\{\omega \in \Omega: |X_n| > \delta\}} |X_n|^r dP \geq \delta^r \int_{\{\omega \in \Omega: |X_n| > \delta\}} dP \geq 0,$$

but

$$\int_{\{\omega \in \Omega: |X_n| > \delta\}} dP = P\{|X_n| > \delta\}.$$

Therefore

$$0 \leq P\{\omega \in \Omega: |X_n| > \delta\} \leq \frac{E|X_n|^r}{\delta^r}. \quad (3.2)$$

The right hand side of 3.2 goes to zero as  $n \rightarrow \infty$  by hypothesis, so that the conclusion of the theorem follows.

If  $r = 2$  in expression 3.2 we obtain Tchebychev's inequality.

It can also be shown that almost sure convergence implies convergence in probability, see for example Tucker (1967). Another important result which we state without proof is the following:

Theorem 2 If  $\underline{X}_n \xrightarrow{P} \underline{X}$  then  $F_{\underline{X}_n} \xrightarrow{d} F_{\underline{X}}$ . For a proof see Tucker (1967).

The last result we shall consider on the subject is given by the following theorem taken from Arnold's (1968) Notes.

Theorem 3 If  $g : R^{k_1} \rightarrow R^{k_2}$  ( $R^{k_1}$  and  $R^{k_2}$  denoting Euclidean  $k_1$  and  $k_2$  - dimensional spaces, respectively) is a continuous function and  $\underline{X}_n \xrightarrow{P} \underline{X}$  where  $\underline{X}_n$  is  $k_1$  - dimensional, then  $g(\underline{X}_n) \xrightarrow{P} g(\underline{X})$  where  $g(\underline{X}_n)$  is  $k_2$  - dimensional.

Proof Let  $\epsilon_1 > 0$  and  $\epsilon_2 > 0$  be given. Let  $A$  be a finite closed  $k_1$  - dimensional rectangle such that  $P(\underline{X} \in A) \geq 1 - \epsilon_2/2$ .  $g$  continuous on its domain implies  $g$  is uniformly continuous on  $A$ . Let  $d(\underline{X}_n, \underline{X})$  denote the distance between the points  $\underline{X}_n$  and  $\underline{X}$  in  $R^{k_1}$  and let  $d[g(\underline{X}_n), g(\underline{X})]$  denote the distance between the points  $g(\underline{X}_n)$  and  $g(\underline{X})$  in  $R^{k_2}$ . Then, there exists  $\delta > 0$  such that if  $\underline{X} \in A$  and  $d(\underline{X}_n, \underline{X}) < \delta$ , we have  $d[g(\underline{X}_n), g(\underline{X})] < \epsilon_1$ . Therefore

$$P\{d[g(\underline{X}_n), g(\underline{X})] \geq \epsilon_1\} \leq P\{\underline{X} \notin A\} + P\{d(\underline{X}_n, \underline{X}) \geq \delta\} \leq \frac{\epsilon_2}{2} + P\{d(\underline{X}_n, \underline{X}) \geq \delta\}.$$

Since by hypothesis  $\underline{X}_n \xrightarrow{P} \underline{X}$ , there exists an  $N$  such that, for  $n \geq N$   $P\{d(\underline{X}_n, \underline{X}) \geq \delta\} \leq \epsilon_2/2$ . By this choice of  $N$  the conclusion of the theorem follows.

An immediate application of the above theorem is the following:

If  $X_n \xrightarrow{P} X$  and  $Y_n \xrightarrow{P} Y$  where both  $X_n$  and  $Y_n$  are 1-dimensional, then  $X_n + Y_n \xrightarrow{P} X + Y$  and  $X_n Y_n \xrightarrow{P} XY$ . The justification of these results is clear.  $X + Y$  and  $XY$  are continuous mappings of the random variables  $X$  and  $Y$ . The theorem is applied with  $g$  being successively defined by  $g(x,y) = x + y$  and  $g(x,y) = xy$ .

Further developments on the above topics are covered in the books by Chung (1968), Tucker (1967), etc., as well as in Arnold's (1968) Notes.

## B. Order in Probability

Let  $\{a_n\}$  be a sequence of real numbers and  $\{r_n\}$  a sequence of positive numbers.

Definition 5 We say  $a_n$  is of order  $o(r_n)$ , or in symbols  $a_n = o(r_n)$ , if

$$\lim_{n \rightarrow \infty} \frac{a_n}{r_n} = 0.$$

Definition 6 We say  $a_n$  is of order  $O(r_n)$ , or in symbols  $a_n = O(r_n)$ , if for some finite real number  $M$

$$\frac{|a_n|}{r_n} \leq M \text{ for all } n.$$

Analogous definitions can be given for sequences of random variables. If  $\{X_n\}$  is a sequence of random variables, we have:

Definition 7 We say  $X_n$  is of order in probability  $o(r_n)$ , or in symbols  $X_n = o_p(r_n)$ , if for each  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\left\{\frac{|X_n|}{r_n} > \epsilon\right\} = 0.$$



Definition 8 We say  $X_n$  is of order in probability  $O(r_n)$ , or in symbols  $X_n = O_p(r_n)$ , if for each  $\epsilon > 0$  there is an  $M_\epsilon$  such that

$$P\{|X_n| > M_\epsilon r_n\} < \epsilon \text{ for all } n.$$

Note: Alternative definitions of order and order in probability assume that  $a_n/r_n$  and  $X_n/r_n$  are bounded for  $n$  greater than some finite  $n_0$ , respectively.

Theorem 4, given below, shows that the algebra of the common order relationships (Definitions 5 and 6, above), extends to the algebra of the order in probability relationships (Definitions 7 and 8, above). Let  $\{X_n\}$  be a sequence of  $k$  dimensional random variables with elements

$$X_n^{(j)} \quad j = 1, 2, \dots, k,$$

and  $g_n(X_n)$  be a sequence of measurable functions. Let  $\{s_n\}$  and  $\{r_n^{(j)}\}$  be  $k+1$  sequences of positive numbers.

Theorem 4 Suppose that

$$\begin{aligned} (a) \quad X_n^{(j)} &= O_p(r_n^{(j)}) \quad j = 1, 2, \dots, t, \\ X_n^{(j)} &= o_p(r_n^{(j)}) \quad j = t+1, t+2, \dots, k, \end{aligned}$$

and that

(b) for any nonrandom sequence  $\{a_n\}$  for which

$$\begin{aligned} a_n^{(j)} &= O(r_n^{(j)}) \quad j = 1, 2, \dots, t, \\ a_n^{(j)} &= o(r_n^{(j)}) \quad j = t+1, t+2, \dots, k, \end{aligned}$$

hold, there exists an  $N$  such that for  $n > N$   $g_n(a_n) = o(s_n)$ .

Then, for  $n > N$   $g_n(X_n) = o_p(s_n)$ . Furthermore, if the last line of hypothesis (b) is replaced by  $g_n(X_n) = o(s_n)$ , the conclusion is  $g_n(X_n) = o_p(s_n)$ .

Theorem 4 was first shown by Mann and Wald (1943), a version of which is also given by Chernoff (1956) and Pratt (1959). The proof given below is contained in Fuller (1970a).

Proof of Theorem 4 Given  $\epsilon > 0$ , define the  $k$  dimensional hypercube,  $A_n$ , by the boundaries

$$\begin{aligned} \pm M^{(j)}_{r_n} & \quad j = 1, 2, \dots, t, \\ \pm M^{(j)}_{r_n} & \quad j = t+1, t+2, \dots, k, \end{aligned}$$

where  $M^{(j)}$  are fixed numbers for  $j = 1, 2, \dots, t$  and  $\lim_{n \rightarrow \infty} M^{(j)}_{r_n} = 0$  for  $j = t+1, t+2, \dots, k$ . The  $M^{(j)}$  and  $M^{(j)}_{r_n}$  are chosen such that

$$\begin{aligned} P\{|X_n^{(j)}| \geq M^{(j)}_{r_n}\} & < \epsilon/k \quad j = 1, 2, \dots, t \\ P\{|X_n^{(j)}| \geq M^{(j)}_{r_n}\} & < \epsilon/k \quad j = t+1, t+2, \dots, k. \end{aligned}$$

Now for  $a \in A_n$  and  $n > N$  there exists an  $M$  such that

$$|g_n(a)|/s_n < M.$$

If not we could form a sequence  $\{a_n\}$  such that  $|g_n(a_n)| > Ms_n$  infinitely often, but this would be contrary to the hypothesis. Hence, for  $X_n \in A_n$

$$g_n(X_n)/s_n < M,$$

and since  $A_n$  was constructed such that

$$P\{X_n \in A_n\} \geq 1 - \epsilon ,$$

it follows that for  $n > N$   $g_n(X_n) = o_p(s_n)$  , showing the first part of the theorem.

By a similar argument for  $n > N$  and  $a \in A_n$

$$|g_n(a)|/s_n < M_n$$

where  $\lim_{n \rightarrow \infty} M_n = 0$  . The second part of the theorem then follows.

Using the definitions and properties of limits it is easily shown that:

(i) If  $a_n = o(f_n)$  and  $b_n = o(g_n)$  then

$$a_n b_n = o(f_n g_n)$$

$$(a_n + b_n) = o\{\text{Max}(f_n, g_n)\} ,$$

(ii) If  $a_n = o(f_n)$  and  $b_n = O(g_n)$  then

$$a_n b_n = o(f_n g_n) .$$

Now, if we have sequences of random variables  $\{X_n\}$  and  $\{Y_n\}$  such that

$$X_n = o_p(f_n) \text{ and } Y_n = o_p(g_n) .$$

It follows by Theorem 4 and the above properties that

$$X_n Y_n = o_p(f_n g_n) , \tag{3.3}$$

$$(X_n + Y_n) = o_p\{\text{Max}(f_n, g_n)\} . \tag{3.4}$$

Similarly, if

$$X_n = O_p(f_n) \text{ and } Y_n = o_p(g_n)$$

then

$$X_n Y_n = O_p(f_n g_n) \quad (3.5)$$

$$(X_n + Y_n) = O_p\{\max(f_n, g_n)\} ; \quad (3.6)$$

etc.

### C. Applications

Let us consider a few examples.

Example 3 Let  $\{X_n\}$  be a sequence of uncorrelated random variables with zero means and constant variances  $\sigma^2$ . Let  $\bar{X}_n = (\sum_{j=1}^n X_j)/n$  ; using Tchebychev's inequality, we obtain

$$P\{|\bar{X}_n| > \sigma \delta\} \leq 1/\delta^2 \text{ for all } n \text{ and } \delta > 0$$

$$P\{|\bar{X}_n| > \sigma\delta/\sqrt{n}\} \leq 1/\delta^2 \text{ for all } n \text{ and } \delta > 0 .$$

In Definition 8 take  $\epsilon = 1/\delta$  ,  $M_\epsilon = \sigma\delta$  and  $r_n = 1/\sqrt{n}$  . Thus,

$$\bar{X}_n = O_p(1/\sqrt{n}) .$$

Example 4 Let  $y = X\beta + e$  , be the general linear regression model, where  $E(e) = 0$  ,  $E(ee') = I\sigma^2$  ,  $X$  an  $n \times p$  matrix of fixed constants such that  $X'X$  is nonsingular,  $\beta$  a  $p \times 1$  vector of unknown parameters and  $y$  the  $n \times 1$  vector of observable random variables. Then we know  $\hat{\beta} = (X'X)^{-1}X'y$  and  $\text{Var}(\hat{\beta}) = (X'X)^{-1}\sigma^2$  . Assume  $1/n X'X = O(1)$ . Then using Example 3 we get, for any component of  $\hat{\beta}$ :

$$P\{|\hat{\beta}_i - \beta_i| > \delta \sqrt{E(\hat{\beta}_i - \beta_i)^2}\} < 1/\delta^2 , \quad i = 1, 2, \dots, p .$$

$$P\{|\hat{\beta}_i - \beta_i| > \delta \sqrt{s_{ii}\sigma^2}\} < 1/\delta^2 ,$$

where  $s^{ii}$  is the element on the  $i$ th row and  $i$ th column of  $(X'X)^{-1}$ . But  $ns^{ii} = O(1)$ . Therefore

$$P\{|\hat{\beta}_i - \beta_i| > \delta \sqrt{ns^{ii}} \sigma/n\} < 1/\delta^2, \text{ for all } n, i = 1, 2, \dots, p.$$

The quantity  $\sigma \delta \sqrt{ns^{ii}} \rightarrow k$  some constant as  $n \rightarrow \infty$ ,  $i = 1, \dots, p$ , for every  $\delta > 0$ . Then  $(\hat{\beta}_i - \beta_i) = O_p(1/\sqrt{n})$ ,  $i = 1, \dots, p$ .

Example 5 Let  $\{X_n\}$  be a sequence of scalar random variables such that

$$X_n = a + O_p(r_n)$$

where  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $g(x)$  be a real valued function with  $s$  continuous derivatives at  $x = a$ . Then by Theorem 4, we have

$$g(X_n) = g(a) + g^{(1)}(a) \cdot (X_n - a) + \dots + \frac{1}{(s-1)!} g^{(s-1)}(a) \cdot (X_n - a)^{s-1} + O_p(r_n^s),$$

where  $g^{(n)}(a)$  is the  $n$ th derivative of  $g(x)$  evaluated at  $x = a$ .

Example 6 Let  $g(\hat{\beta})$  be a real valued differentiable function of the  $p$ -variate vector  $\hat{\beta}$ , the estimator of  $\beta$  in Example 4. It is not difficult to establish the following results:

$$E\{g(\hat{\beta})\} = g(\beta) + O_p(1/n)$$

$$E\{g(\hat{\beta}) - g(\beta)\}^2 = \sum_{i=1}^p \sum_{j=1}^p \frac{\partial g}{\partial \hat{\beta}_i} \frac{\partial g}{\partial \hat{\beta}_j} \bigg|_{\hat{\beta}=\beta} s^{ij} \sigma^2 + O_p(1/n^{3/2}).$$

Example 7

Meier (1953) gives the following result: If

$z_1, \dots, z_p$  are independently distributed random variables with density functions

$$f_{n_i}(z_i) = \frac{\left(\frac{1}{2}n_i\right)^{\frac{1}{2}n_i}}{\Gamma\left(\frac{1}{2}n_i\right)} z_i^{\left(\frac{1}{2}n_i\right)-1} e^{-\frac{1}{2}n_i z_i} \quad 0 \leq z_i < \infty$$

and  $R(z_1, \dots, z_k)$  is a rational function with no singularities for

$0 < z_1, \dots, z_k < \infty$ , then  $E[R(z_1, \dots, z_k)]$  can be expanded in an asymptotic series in the  $1/n_i$ :

$$E[R(z_1, \dots, z_k)] = R(1, \dots, 1) + \sum_{i=1}^k \frac{1}{n_i} \frac{\partial^2 R}{\partial z_i^2} \Big|_{1 \dots 1} + O\left(\sum_{i=1}^k \frac{1}{n_i^2}\right).$$

It is easily shown that a better approximation to  $E[R(z_1, \dots, z_k)]$  is obtained by taking

$$E[R(z_1, \dots, z_k)] = R(1, \dots, 1) + \sum_{i=1}^k \frac{1}{n_i} \frac{\partial^2 R}{\partial z_i^2} \Big|_{1 \dots 1} + \sum_{i=1}^k \frac{4}{3n_i^2} \frac{\partial^3 R}{\partial z_i^3} \Big|_{1 \dots 1} + O\left(\sum_{i=1}^k \frac{1}{n_i^3}\right).$$

#### IV. ESTIMATION OF THE SOLUTION OF LINEAR EQUATIONS WITH RANDOM COEFFICIENTS

We discuss in this chapter the general errors in variables model. Starting with a simple estimator, an estimator unbiased to  $O_p(1/n)$  is presented. The efficiency of the unbiased estimator with respect to the initial one is investigated. A method to assure finiteness of the first moments of the modified estimator is also presented. An application to the estimation in a simultaneous equations problem is also covered.

##### A. Nearly Unbiased Estimators

###### 1. Introduction

Fuller (1968), has considered the estimation of the vector  $\beta$  in the model

$$\begin{aligned} y &= x\beta \\ Y &= y + e \\ X &= x + u \end{aligned} \tag{4.1}$$

where  $X$ ,  $x$  and  $u$  are  $n \times k$  matrices and  $Y$ ,  $y$  and  $e$  are  $n \times 1$  vectors.  $x$  may be either a fixed matrix or a random sample from a  $k$ -variate population with nonsingular covariance matrix. The  $n \times (k+1)$  matrix of errors

$$(u, e) = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1k} & e_1 \\ u_{21} & u_{22} & \cdots & u_{2k} & e_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ u_{n1} & u_{n2} & \cdots & u_{nk} & e_n \end{bmatrix},$$

is such that the vectors  $(u_{i1}, u_{i2}, \dots, u_{ik}, e_i)'$ ,  $i = 1, 2, \dots, n$ , are independently distributed with the same multivariate normal distribution

$$N_{(k+1)} \left[ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} & \psi_1 \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2k} & \psi_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk} & \psi_k \\ \psi_1 & \psi_2 & \dots & \psi_k & \sigma_e^2 \end{pmatrix} \right]$$

$$= N_{(k+1)} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \psi \\ \psi' & \sigma_e^2 \end{pmatrix} \right].$$

Consider a system of equations of the form

$$H\beta = N \quad (4.2)$$

where  $H$  is a fixed positive definite  $k \times k$  matrix and  $N$  a fixed  $k \times 1$  vector. It is assumed that there are available estimators  $\hat{H}$  and  $\hat{N}$  satisfying

$$\begin{aligned} \hat{H} &= H + a \\ \hat{N} &= N + b \end{aligned} \quad (4.3)$$

where the elements of  $a$ ,  $(k \times k)$ , and  $b$ ,  $(k \times 1)$ , are  $O_p(1/\sqrt{n})$  and have zero expectation. Consider the estimator  $\hat{\beta} = \hat{H}^{-1} \hat{N}$  which can be expressed in terms of the following expansion:



$$\begin{aligned}
\hat{\beta} &= \hat{H}^{-1} \hat{N} \\
&= (H+a)^{-1} (N+b) \\
&= [I - H^{-1}a + H^{-1}aH^{-1}a - \dots] H^{-1} (N+b) \\
\hat{\beta} &= \{H^{-1}N - H^{-1}aH^{-1}N + H^{-1}aH^{-1}aH^{-1}N + H^{-1}b - H^{-1}aH^{-1}b\} \\
&\quad + O_p(n^{-3/2})
\end{aligned} \tag{4.4}$$

Thus, by neglecting terms whose order in probability is less than  $1/n$ , the terms within the brackets on the right hand side of 4.4 will converge in distribution to the limiting distribution of  $\hat{\beta}$  as  $n \rightarrow \infty$ . To this order of approximation the asymptotic expectation of  $\hat{\beta}$  is given by

$$\bar{E}(\hat{\beta}) = \beta + E(H^{-1}aH^{-1}a\beta) - E(H^{-1}aH^{-1}b) \tag{4.5}$$

where in 4.5,  $H^{-1}N = \beta$ .

Assume that in the construction of an estimator for  $\beta$  estimators of  $\Sigma$  and  $\psi$  are available, that these estimators are obtained from a single sample of normal observations. That is,  $\hat{\Sigma} = S$  has a Wishart distribution with  $d$  degrees of freedom. It is also assumed that  $S = \hat{\Sigma}$  and  $R = \hat{\psi}$  are independent of  $x$ ,  $u$  and  $e$  and that  $\sigma_e^2$  is not necessarily known.

Now, in terms of the variables in Model 4.1

$$\begin{aligned}
H &= \frac{1}{n} x'x, \\
N &= \frac{1}{n} x'y,
\end{aligned} \tag{4.6}$$

where  $y = x\beta$ ;

$$\begin{aligned}
\hat{H} &= \frac{1}{n} X'X - S, \\
\hat{N} &= \frac{1}{n} X'Y - R.
\end{aligned} \tag{4.7}$$

Further

$$\begin{aligned} a &= \frac{1}{n} (x'u + u'x + u'u) - S \\ b &= \frac{1}{n} (u'y + x'e + u'e) - R \end{aligned} \quad (4.8)$$

Define the difference

$$\bar{E}(\hat{\beta}) - \beta = E(H^{-1}aH^{-1}a\beta) - E(H^{-1}aH^{-1}b)$$

as the asymptotic bias of the estimator  $\hat{\beta}$ . Under the above description of the estimation problem, Fuller (1968) has obtained the following expression for the asymptotic bias of  $\hat{\beta}$

$$\begin{aligned} \bar{E}(\hat{\beta}) - \beta &= \frac{1}{n} H^{-1} \{ [(k+1+\alpha \text{tr} H^{-1})\bar{Z} + \alpha \bar{Z} H^{-1}] \beta \\ &\quad - [(k+1+\alpha \text{tr} H^{-1})\bar{\psi} + \alpha \bar{\psi} H^{-1}] \} \end{aligned} \quad (4.9)$$

where  $\alpha = (1 + n/d)$ . Based on the bias he constructs the estimator

$$\tilde{\beta} = (\hat{H} + \hat{G})^{-1}(\hat{N} + \hat{D}) \quad , \quad (4.10)$$

where

$$\begin{aligned} \hat{G} &= \frac{1}{n} (k+1+\alpha \text{tr} \hat{H}^{-1})S + \left(\frac{1}{n} + \frac{1}{d}\right) \hat{S} \hat{H}^{-1} S \\ \hat{D} &= \frac{1}{n} (k+1+\alpha \text{tr} \hat{H}^{-1})R + \left(\frac{1}{n} + \frac{1}{d}\right) \hat{S} \hat{H}^{-1} R \end{aligned} \quad (4.11)$$

which is asymptotically unbiased for  $\beta$  to  $O_p(1/n)$ .

Fuller (1968) gives also the following result. The mean square error of  $\tilde{\beta}$  to  $O_p(1/n)$  is given by

$$\begin{aligned} \text{M.S.E.}(\tilde{\beta}) &= \frac{1}{n} H^{-1} (\sigma_e^2 - 2\beta' \psi + \beta' \psi \psi' \beta) + \left(\frac{1}{n} + \frac{1}{d}\right) (\sigma_e^2 - 2\beta' \psi + \beta' \psi \psi' \beta) \cdot H^{-1} \psi \psi' H^{-1} \\ &+ \left(\frac{1}{n} + \frac{1}{d}\right) \cdot H^{-1} (\psi \psi' \beta) (\psi \psi' \beta)' H^{-1} . \end{aligned} \quad (4.12)$$

It is easily shown that to  $O_p(1/n)$   $\hat{\beta} = \hat{H}^{-1} \hat{N}$  and  $\tilde{\beta}$  given by 4.10 have the same mean square error. In the following section we propose to obtain the general conditions under which, to  $O_p(1/n^2)$ ,  $\tilde{\beta}$  has smaller mean square error than  $\hat{\beta}$ .

## 2. Relative efficiency of $\hat{\beta}$ and $\tilde{\beta}$

Rather than evaluate both mean square errors to order  $(1/n^2)$  we shall consider the difference

$$\bar{E}\{(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\} - \bar{E}\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\} .$$

Setting

$$\hat{u} = \hat{D} - \hat{G}\beta ,$$

and

$$\bar{u} = D - G\beta = E\{aH^{-1}w\}$$

where  $w = b - a\beta$ . We observe that

$$\begin{aligned} \bar{u} &= O(1/n) \\ w &= O_p(1/\sqrt{n}) . \end{aligned}$$

We have successively

$$\begin{aligned} \tilde{\beta} &= (\hat{H} + \hat{G})^{-1}(\hat{N} + \hat{D}) \\ &= (H + a + \hat{G})^{-1}(N + b + \hat{D}) \\ &= [I + H^{-1}(a + \hat{G})]^{-1} H^{-1} [N + (b + \hat{D})] \\ &= [H^{-1} - H^{-1}(a + \hat{G})H^{-1} + H^{-1}(a + \hat{G})H^{-1}(a + \hat{G})H^{-1} - \dots][N + (b + \hat{D})] \end{aligned}$$

$$\begin{aligned}
&= H^{-1}N^{-1}H^{-1}(a+\hat{G})H^{-1}N + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}N + H^{-1}(b+\hat{D}) \\
&\quad - H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + \dots
\end{aligned}$$

$$\begin{aligned}
\tilde{\beta} - \beta &= - H^{-1}(a+\hat{G})\beta + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})\beta + H^{-1}(b+\hat{D}) - H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) \\
&\quad + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + \dots \\
&= - H^{-1}[(a+\hat{G})\beta - (b+\hat{D})] + H^{-1}(a+\hat{G})H^{-1}[(a+\hat{G})\beta - (b+\hat{D})] \\
&\quad + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + \dots \\
&= - H^{-1}[(a\beta - b) + (\hat{G}\beta - \hat{D})] + H^{-1}(a+\hat{G})H^{-1}[(a\beta - b) + (\hat{G}\beta - \hat{D})] \\
&\quad + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + \dots
\end{aligned}$$

If we recall that  $b - a\beta = w$  and  $\hat{D} - \hat{G}\beta = \hat{u}$ , we can write:

$$\tilde{\beta} - \beta = H^{-1}(w+\hat{u}) - H^{-1}(a+\hat{G})H^{-1} + H^{-1}(a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D}) + \dots$$

Therefore, to order in probability  $(1/n^2)$

$$\begin{aligned}
(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' &= H^{-1}\{(w+\hat{u})(w+\hat{u})' - (w+\hat{u})(w+\hat{u})'H^{-1}(a+\hat{G})' \\
&\quad + (w+\hat{u})(b+\hat{D})'H^{-1}(a+\hat{G})'H^{-1}(a+\hat{G})' - (a+\hat{G})H^{-1}(w+\hat{u})(w+\hat{u})' \\
&\quad + (a+\hat{G})H^{-1}(w+\hat{u})(w+\hat{u})'H^{-1}(a+\hat{G})' + (a+\hat{G})H^{-1}(a+\hat{G})H^{-1}(b+\hat{D})(w+\hat{u})'\}H^{-1} \\
&= H^{-1}\{(w+\hat{u})(w+\hat{u})' - (w+\hat{u})(w+\hat{u})'H^{-1}(a+\hat{G})' + wb'H^{-1}a'H^{-1}a' \\
&\quad - (a+\hat{G})H^{-1}(w+\hat{u})(w+\hat{u})' + aH^{-1}ww'H^{-1}a' + aH^{-1}aH^{-1}bw'\}H^{-1} \\
&= H^{-1}\{ww' + w\hat{u}' + \hat{u}w' + \bar{u}\bar{u}' - (ww' + w\bar{u}' + \bar{u}w')H^{-1}a' \\
&\quad - ww'H^{-1}G' + wb'H^{-1}a'H^{-1}a' - aH^{-1}(ww' + w\bar{u}' + \bar{u}w') \\
&\quad - GH^{-1}ww' + aH^{-1}aH^{-1}bw' + aH^{-1}ww'H^{-1}a'\}H^{-1} \\
&= H^{-1}\{ww' + w\hat{u}' + \hat{u}w' + \bar{u}\bar{u}' - ww'H^{-1}a' - w\bar{u}'H^{-1}a' - \bar{u}w'H^{-1}a' \\
&\quad - ww'H^{-1}G' + wb'H^{-1}a'H^{-1}a' - aH^{-1}ww' - aH^{-1}w\bar{u}' - aH^{-1}\bar{u}w' \\
&\quad - GH^{-1}ww' + aH^{-1}aH^{-1}bw' + aH^{-1}ww'H^{-1}a'\}H^{-1}
\end{aligned}$$

$$\begin{aligned}
(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)' &= (\hat{\beta}-\beta)(\hat{\beta}-\beta)' \\
&+ H^{-1} \hat{w} \hat{u}' H^{-1'} + H^{-1} \hat{u} \hat{w}' H^{-1'} \\
&+ H^{-1} \bar{u} \bar{u}' H^{-1'} \\
&- H^{-1} \bar{w} \bar{w}' H^{-1'} G' H^{-1'} - H^{-1} G H^{-1} \bar{w} \bar{w}' H^{-1'} \\
&- H^{-1} \bar{u} \bar{w}' H^{-1'} a' H^{-1'} - H^{-1} a H^{-1} \bar{w} \bar{u}' H^{-1'} \\
&- H^{-1} \bar{w} \bar{u}' H^{-1'} a' H^{-1'} - H^{-1} a H^{-1} \bar{u} \bar{w}' H^{-1'}
\end{aligned} \tag{4.13}$$

after we observe that to  $O_p(1/n^2)$

$$\begin{aligned}
(\hat{\beta}-\beta)(\hat{\beta}-\beta)' &= H^{-1} \bar{w} \bar{w}' H^{-1'} - H^{-1} \bar{w} \bar{w}' H^{-1'} a' H^{-1'} + H^{-1} \bar{w} b' H^{-1'} a' H^{-1'} a' H^{-1'} \\
&- H^{-1} a H^{-1} \bar{w} \bar{w}' H^{-1'} + H^{-1} a H^{-1} \bar{w} \bar{w}' H^{-1'} a' H^{-1'} + H^{-1} a H^{-1} a H^{-1} b w' H^{-1'} .
\end{aligned}$$

We shall obtain the expected value of the difference

$$(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)' - (\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$

term by term.

Calculation of  $E\{H^{-1} \hat{u} \hat{w}' H^{-1'}\}$  : Using expression 4.9, we obtain

$$\bar{u} = E\{a H^{-1} w\} = \frac{1}{n} \{[k+1+\alpha \text{tr}(H^{-1} \Sigma)] + \alpha H^{-1}\}(\Psi - \Sigma \beta)$$

also, from expressions 4.11,

$$\bar{u} = \hat{D} - \hat{G}\beta = \frac{1}{n} \{[k+1+\alpha \text{tr}(\hat{H}^{-1} \Sigma)] + \alpha \hat{H}^{-1}\}(\Psi - \Sigma \beta)$$

Now  $\hat{H} = H + a$ , therefore working with the above expression we obtain

$$\begin{aligned}
\hat{u} &= \frac{1}{n} \{[k+1+\alpha \text{tr}(H+a)^{-1} \Sigma] + \alpha (H+a)^{-1}\}(\hat{\Psi} - \hat{\Sigma} \beta) \\
&= \frac{1}{n} \{[k+1+\alpha \text{tr}(H^{-1} - H^{-1} a H^{-1} + \dots) \Sigma] + \alpha (H^{-1} - H^{-1} a H^{-1} + \dots)\}(\hat{\Psi} - \hat{\Sigma} \beta) + \dots \\
&= \frac{1}{n} \{[k+1+\alpha \text{tr}(H^{-1} \Sigma)] + \alpha H^{-1}\}(\hat{\Psi} - \hat{\Sigma} \beta) - \alpha [\text{tr}(H^{-1} a H^{-1} \Sigma) \\
&\quad + H^{-1} a H^{-1} \Sigma] (\Psi - \Sigma \beta) + \dots
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \{ [k+1+\alpha \text{tr}(\mathbf{H}^{-1}\mathbf{Z}) + \alpha \mathbf{Z}\mathbf{H}^{-1}] (\hat{\Psi}-\hat{\mathbf{Z}}\beta) + \alpha [\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z}) + (\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}] \\
&\quad (\hat{\Psi}-\hat{\mathbf{Z}}\beta) - \alpha [\text{tr}(\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}\mathbf{Z}) + \mathbf{Z}\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}] (\hat{\Psi}-\hat{\mathbf{Z}}\beta) \} + \dots \\
&= \frac{1}{n} \{ [k+1+\alpha \text{tr}(\mathbf{H}^{-1}\mathbf{Z}) + \alpha \mathbf{Z}\mathbf{H}^{-1}] (\hat{\Psi}-\hat{\mathbf{Z}}\beta) + \alpha [\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z}) \\
&\quad + (\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}] (\hat{\Psi}-\mathbf{Z}\beta) - \alpha [\text{tr}(\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}\mathbf{Z}) + \mathbf{Z}\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}] (\hat{\Psi}-\mathbf{Z}\beta) \} + \dots
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\mathbf{u}} - \mathbf{u} &= \frac{1}{n} [k+1+\alpha \text{tr}(\mathbf{H}^{-1}\mathbf{Z}) + \alpha \mathbf{Z}\mathbf{H}^{-1}] [(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta] + \frac{\alpha}{n} [\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z}) \\
&\quad + (\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}] (\hat{\Psi}-\mathbf{Z}\beta) - \frac{\alpha}{n} [\text{tr}(\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}\mathbf{Z}) + \mathbf{Z}\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}] (\hat{\Psi}-\mathbf{Z}\beta) + \dots
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\{\mathbf{H}^{-1}\hat{\mathbf{u}}\mathbf{w}'\mathbf{H}^{-1}\} &= \mathbf{E}\{\mathbf{H}^{-1}(\hat{\mathbf{u}}-\mathbf{u})\mathbf{w}'\mathbf{H}^{-1}\} \\
&= \mathbf{H}^{-1}\mathbf{E}\left\{\frac{1}{n}[k+1+\alpha \text{tr}(\mathbf{H}^{-1}\mathbf{Z}) + \alpha \mathbf{Z}\mathbf{H}^{-1}][(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta]\mathbf{w}'\right. \\
&\quad + \frac{\alpha}{n}[\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z}) + (\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}](\hat{\Psi}-\mathbf{Z}\beta)\mathbf{w}' - \frac{\alpha}{n}[\text{tr}(\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}\mathbf{Z}) \\
&\quad + \mathbf{Z}\mathbf{H}^{-1}\mathbf{a}\mathbf{H}^{-1}](\hat{\Psi}-\mathbf{Z}\beta)\mathbf{w}'\} \mathbf{H}^{-1}. \tag{4.14}
\end{aligned}$$

It can be shown that

$$\begin{aligned}
\mathbf{E}\{[(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta]\mathbf{w}'\} &= -\mathbf{E}\{[(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta][(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta]'\} \\
&= -\frac{1}{d} [(\sigma_e^2 - 2\Psi'\beta + \beta'\mathbf{Z}\beta) \cdot \mathbf{Z} + (\Psi-\mathbf{Z}\beta)(\Psi-\mathbf{Z}\beta)'] \tag{4.15}
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\{[\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z})](\hat{\Psi}-\mathbf{Z}\beta)\mathbf{w}'\} &= -\mathbf{E}\{[\text{tr}\mathbf{H}^{-1}(\hat{\mathbf{Z}}-\mathbf{Z})](\hat{\Psi}-\mathbf{Z}\beta)[(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta]\} \\
&= -\frac{2}{d} (\Psi-\mathbf{Z}\beta)(\Psi-\mathbf{Z}\beta)'\mathbf{H}^{-1}\mathbf{Z} \tag{4.16}
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\{(\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}(\hat{\Psi}-\mathbf{Z}\beta)\mathbf{w}'\} &= -\mathbf{E}\{(\hat{\mathbf{Z}}-\mathbf{Z})\mathbf{H}^{-1}(\hat{\Psi}-\mathbf{Z}\beta)[(\hat{\Psi}-\Psi) - (\hat{\mathbf{Z}}-\mathbf{Z})\beta]\} \\
&= -\frac{1}{d} [(\Psi-\mathbf{Z}\beta)'\mathbf{H}^{-1}(\Psi-\mathbf{Z}\beta) \cdot \mathbf{Z} + (\Psi-\mathbf{Z}\beta)(\Psi-\mathbf{Z}\beta)'\mathbf{H}^{-1}\mathbf{Z}] \tag{4.17}
\end{aligned}$$

$$\begin{aligned}
E\{[\text{tr}(H^{-1}aH^{-1}Z)](\psi-Z\beta)w'\} &= E\{[\text{tr}(H^{-1}ZH^{-1}a)](\psi-Z\beta)(b'-\beta'a')\} \\
&= \frac{2}{n}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}Z + 2\left(\frac{1}{n} + \frac{1}{d}\right)(\psi-Z\beta)(\psi-Z\beta)'H^{-1}ZH^{-1}Z
\end{aligned} \tag{4.18}$$

$$\begin{aligned}
E\{ZH^{-1}aH^{-1}(\psi-Z\beta)w'\} &= E\{ZH^{-1}aH^{-1}(\psi-Z\beta)(b'-\beta'a')\} \\
&= \frac{1}{n}\{(\psi-Z\beta)'H^{-1}(\psi-Z\beta) \cdot ZH^{-1}Z + ZH^{-1}(\psi-Z\beta)(\psi-Z\beta)'\} \\
&\quad + \left(\frac{1}{n} + \frac{1}{d}\right)\{ZH^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}Z + (\psi-Z\beta)'H^{-1}(\psi-Z\beta) \cdot ZH^{-1}Z\}
\end{aligned} \tag{4.19}$$

where we have used the following results given by Fuller (1968).

$$\begin{aligned}
\text{Cov}(a_{ij}, a_{mt}) &= \frac{1}{n}(h_{im}\sigma_{jt} + h_{it}\sigma_{jm} + h_{jm}\sigma_{it} + h_{jt}\sigma_{im}) + \left(\frac{1}{n} + \frac{1}{d}\right)(\sigma_{it}\sigma_{jm} + \sigma_{im}\sigma_{jt}), \\
\text{Cov}(a_{ij}, b_m) &= \frac{1}{n}(n_{i\cdot}\sigma_{mj} + n_{j\cdot}\sigma_{im} + h_{mi}\sigma_{ej} + h_{mj}\sigma_{ei}) + \left(\frac{1}{n} + \frac{1}{d}\right)(\sigma_{ei}\sigma_{jm} + \sigma_{im}\sigma_{ej}).
\end{aligned}$$

Placing the values given by 4.15 through 4.19 in 4.14, we obtain

$$\begin{aligned}
E\{H^{-1}\hat{u}w'H^{-1}\} &= -\frac{1}{nd} [k+1+\alpha\text{tr}(ZH^{-1})][(\sigma_e^2-2\psi'\beta+\beta'Z\beta) \cdot H^{-1}ZH^{-1} + H^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}] \\
&\quad - \frac{1}{d}\left(\frac{1}{n} + \frac{1}{d}\right) \cdot (\sigma_e^2-2\psi'\beta+\beta'Z\beta)H^{-1}ZH^{-1}ZH^{-1} - \frac{1}{d}\left(\frac{1}{n} + \frac{1}{d}\right) \\
&\quad \cdot H^{-1}ZH^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1} \\
&\quad - \frac{1}{n}\left(\frac{1}{n} + \frac{1}{d}\right) \cdot (\psi-Z\beta)'H^{-1}(\psi-Z\beta) \cdot H^{-1}ZH^{-1} - \frac{1}{n}\left(\frac{1}{n} + \frac{1}{d}\right) \\
&\quad \cdot H^{-1}ZH^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1} \\
&\quad - \left(\frac{1}{n} + \frac{1}{d}\right)^2 \cdot H^{-1}ZH^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}ZH^{-1} - \left(\frac{1}{n} + \frac{1}{d}\right)^2(\psi-Z\beta)'H^{-1}(\psi-Z\beta) \\
&\quad \cdot H^{-1}ZH^{-1}ZH^{-1} \\
&\quad - \frac{2}{n}\left(\frac{1}{n} + \frac{1}{d}\right) \cdot H^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}ZH^{-1} - 2\left(\frac{1}{n} + \frac{1}{d}\right)^2 \\
&\quad \cdot H^{-1}(\psi-Z\beta)(\psi-Z\beta)'H^{-1}ZH^{-1}ZH^{-1}.
\end{aligned}$$

$$\begin{aligned}
& - \frac{1}{d} \left( \frac{1}{n} + \frac{1}{d} \right) \cdot (\psi - \gamma \beta)' H^{-1} (\psi - \gamma \beta) \cdot H^{-1} \gamma H^{-1} - \frac{3}{d} \left( \frac{1}{n} + \frac{1}{d} \right) \\
& \quad \cdot H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \gamma H^{-1} .
\end{aligned} \tag{4.20}$$

Also,  $E\{H^{-1} \hat{w} u' H^{-1}\} = (E\{H^{-1} \hat{u} u' H^{-1}\})'$ .

Calculation of  $E\{H^{-1} \hat{u} u' H^{-1}\}$  : We immediately obtain

$$\begin{aligned}
E\{H^{-1} \hat{u} u' H^{-1}\} &= H^{-1} \left\{ \frac{1}{n} [k+1 + \alpha \text{tr}(\gamma H^{-1}) + \alpha \gamma H^{-1}] (\psi - \gamma \beta) (\psi - \gamma \beta)' [k+1 + \alpha \text{tr}(H^{-1} \gamma)] \right. \\
&\quad \left. + \alpha H^{-1} \gamma \right\} \frac{1}{n} H^{-1} = H^{-1} \hat{u} u' H^{-1} \\
&= \frac{1}{n^2} [k+1 + \alpha \text{tr}(\gamma H^{-1})]^2 \cdot H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} + \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+1 + \alpha \text{tr}(\gamma H^{-1})] \\
&\quad \cdot H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \gamma H^{-1} \\
&\quad + \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+1 + \alpha \text{tr}(H^{-1} \gamma)] \cdot H^{-1} \gamma H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \\
&\quad + \left( \frac{1}{n} + \frac{1}{d} \right)^2 \cdot H^{-1} \gamma H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \gamma H^{-1} .
\end{aligned}$$

Calculation of  $E\{H^{-1} \hat{w} w' H^{-1} G' H^{-1}\}$  : It can be seen that

$$\begin{aligned}
E\{H^{-1} \hat{w} w' H^{-1} G' H^{-1}\} &= \{H^{-1} E(w w') H^{-1}\} E(a H^{-1} a) H^{-1} \\
&= \frac{1}{n^2} (\sigma_e^2 - 2\psi' \beta - \beta' \gamma \beta) [k+2 + 2\alpha \text{tr}(\gamma H^{-1})] \cdot H^{-1} \gamma H^{-1} \\
&\quad + \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) (\sigma_e^2 - 2\psi' \beta + \beta' \gamma \beta) [k+3 + 2\alpha \text{tr}(\gamma H^{-1})] \cdot H^{-1} \gamma H^{-1} \gamma H^{-1} \\
&\quad + \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+2 + \alpha \text{tr}(\gamma H^{-1})] \cdot H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \gamma H^{-1} \\
&\quad + \frac{1}{n^2} (\sigma_e^2 - 2\psi' \beta + \beta' \gamma \beta) \text{tr}(\gamma H^{-1}) \cdot H^{-1} \\
&\quad + \left( \frac{1}{n} + \frac{1}{d} \right)^2 (\sigma_e^2 - 2\psi' \beta + \beta' \gamma \beta) \cdot H^{-1} \gamma H^{-1} \gamma H^{-1} \gamma H^{-1} \\
&\quad + \left( \frac{1}{n} + \frac{1}{d} \right)^2 \cdot H^{-1} (\psi - \gamma \beta) (\psi - \gamma \beta)' H^{-1} \gamma H^{-1} \gamma H^{-1} .
\end{aligned}$$

Also  $E\{H^{-1} \hat{G} H^{-1} \hat{w} w' H^{-1}\} = (E\{H^{-1} \hat{w} w' H^{-1} G' H^{-1}\})'$ .



Calculation of  $E\{H^{-1}aH^{-1}\bar{w}u'H^{-1}\}$  : In this case

$$E\{H^{-1}aH^{-1}\bar{w}u'H^{-1}\} = H^{-1}E(aH^{-1}\bar{w})u'H^{-1} = H^{-1}\bar{u}u'H^{-1} .$$

Also  $E\{H^{-1}\bar{u}w'H^{-1}a'H^{-1}\} = (E\{H^{-1}aH^{-1}\bar{w}u'H^{-1}\})'$  .

Calculation of  $E\{H^{-1}\bar{w}u'H^{-1}a'H^{-1}\}$  : Let  $r = H^{-1}\bar{w}$  ,  $s = H^{-1}a$  and  $\delta = H^{-1}\bar{u}$  , then

$$E\{H^{-1}\bar{w}u'H^{-1}a'H^{-1}\} = E\{r\delta's'\} .$$

It is easily shown that the element on the  $j$ th row,  $m$ th column of the above matrix of expected values, is

$$E\{r\delta's'\}_{jm} = \sum_k \delta_k \sum_s \sum_q h^{js} h^{qm} \{E(b_{s kq} - \sum_t \beta_t E(a_{st} a_{kq}))\} .$$

After some algebra we find

$$\begin{aligned} E\{H^{-1}\bar{w}u'H^{-1}a'H^{-1}\} &= \frac{1}{n^2} [k+1+\alpha \text{tr}(\bar{y}H^{-1})] \cdot H^{-1}(\psi-\bar{y}\beta)(\psi-\bar{y}\beta)'H^{-1} \\ &+ \frac{1}{n^2} [k+1+\alpha \text{tr}(\bar{y}H^{-1})] \cdot (\psi-\bar{y}\beta)'H^{-1}(\psi-\bar{y}\beta) \cdot H^{-1} \\ &+ \frac{1}{n}(\frac{1}{n} + \frac{1}{d})[k+1+\alpha \text{tr}(\bar{y}H^{-1})] \cdot (\psi-\bar{y}\beta)'H^{-1}(\psi-\Sigma\beta) \cdot H^{-1}\bar{y}H^{-1} \\ &+ \frac{1}{n}(\frac{1}{n} + \frac{1}{d})[k+1+\alpha \text{tr}(\bar{y}H^{-1})] \cdot H^{-1}\bar{y}H^{-1}(\psi-\bar{y}\beta)(\psi-\bar{y}\beta)'H^{-1} \\ &+ \frac{1}{n}(\frac{1}{n} + \frac{1}{d}) \cdot H^{-1}\bar{y}H^{-1}(\psi-\bar{y}\beta)(\psi-\bar{y}\beta)'H^{-1} \\ &+ \frac{1}{n}(\frac{1}{n} + \frac{1}{d}) \cdot (\psi-\bar{y}\beta)'H^{-1}\bar{y}H^{-1}(\psi-\bar{y}\beta) \cdot H^{-1} \\ &+ (\frac{1}{n} + \frac{1}{d})^2 \cdot (\psi-\bar{y}\beta)'H^{-1}\bar{y}H^{-1}(\psi-\bar{y}\beta) \cdot H^{-1}\bar{y}H^{-1} \\ &+ (\frac{1}{n} + \frac{1}{d})^2 \cdot H^{-1}\bar{y}H^{-1}\bar{y}H^{-1}(\psi-\bar{y}\beta)(\psi-\bar{y}\beta)'H^{-1} \end{aligned}$$

$$\text{Also, } E\{H^{-1}aH^{-1}\bar{u}w'H^{-1}\} = (E\{H^{-1}\bar{w}u'H^{-1}a'H^{-1}\})'.$$

Placing the above results in 4.13, we finally obtain

$$\begin{aligned} & E\{(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'\} - E\{(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\} \\ &= -\frac{2}{nd}[k+1+\alpha\text{tr}(\Sigma H^{-1})][\sigma_w^2 \cdot H^{-1}\Sigma H^{-1} + H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}] \\ &\quad - \frac{2}{d}(\frac{1}{n} + \frac{1}{d}) \cdot \sigma_w^2 \cdot H^{-1}\Sigma H^{-1}\Sigma H^{-1} \\ &\quad - \frac{1}{d}(\frac{1}{n} + \frac{1}{d}) \cdot [H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} + H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1}] \\ &\quad - \frac{2}{n}(\frac{1}{n} + \frac{1}{d}) \cdot (\Psi-\Sigma\beta)'H^{-1}(\Psi-\Sigma\beta) \cdot H^{-1}\Sigma H^{-1} \\ &\quad - \frac{3}{n}(\frac{1}{n} + \frac{1}{d}) \cdot [H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} + H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1}] \\ &\quad - 2(\frac{1}{n} + \frac{1}{d})^2 \cdot H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1} \\ &\quad - 2(\frac{1}{n} + \frac{1}{d})^2 \cdot (\Psi-\Sigma\beta)'H^{-1}(\Psi-\Sigma\beta) \cdot H^{-1}\Sigma H^{-1}\Sigma H^{-1} \\ &\quad - 2(\frac{1}{n} + \frac{1}{d})^2 \cdot [H^{-1}\Sigma H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} \\ &\quad + H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1}\Sigma H^{-1}] \\ &\quad - \frac{2}{d}(\frac{1}{n} + \frac{1}{d}) \cdot (\Psi-\Sigma\beta)'H^{-1}(\Psi-\Sigma\beta) \cdot H^{-1}\Sigma H^{-1} \\ &\quad - \frac{3}{d}(\frac{1}{n} + \frac{1}{d})[H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} + H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1}] \\ &\quad - \frac{1}{n^2}[k+1+\alpha\text{tr}(\Sigma H^{-1})]^2 \cdot H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} \\ &\quad - \frac{1}{n}(\frac{1}{n} + \frac{1}{d})[k+1+\alpha\text{tr}(\Sigma H^{-1})] \cdot H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1}\Sigma H^{-1} \\ &\quad - \frac{1}{n}(\frac{1}{n} + \frac{1}{d})[k+1+\alpha\text{tr}(\Sigma H^{-1})] \cdot H^{-1}\Sigma H^{-1}(\Psi-\Sigma\beta)(\Psi-\Sigma\beta)'H^{-1} \end{aligned}$$

$$\begin{aligned}
& - \frac{1}{n} + \frac{1}{d} \sigma^2 \cdot H^{-1} X_H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} X_H^{-1} \\
& - \frac{2\sigma^2}{n^2} [k+2+2 \alpha \text{tr}(X_H^{-1})] H^{-1} X_H^{-1} - \frac{2}{n} \left( \frac{1}{n} + \frac{1}{d} \right) \sigma^2 [k+3+\alpha t_r(X_H^{-1})] \cdot H^{-1} X_H^{-1} \Sigma^{-1} \\
& - \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+2+\alpha \text{tr}(X_H^{-1})] \cdot [H^{-1} X_H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} + H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} X_H^{-1}] \\
& - \frac{2\sigma^2}{n^2} \text{tr}(X_H^{-1}) \cdot H^{-1} - \frac{2}{n} \left( \frac{1}{n} + \frac{1}{d} \right) t_r(X_H^{-1}) \cdot H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} \\
& - 2 \left( \frac{1}{n} + \frac{1}{d} \right) \sigma^2 \cdot H^{-1} X_H^{-1} X_H^{-1} X_H^{-1} \\
& - \left( \frac{1}{n} + \frac{1}{d} \right) \sigma^2 [H^{-1} X_H^{-1} X_H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} + H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} X_H^{-1} X_H^{-1}] \\
& - \frac{2}{n^2} [k+1+\alpha \text{tr}(X_H^{-1})] \cdot H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} \\
& - \frac{2}{n^2} [k+1+\alpha \text{tr}(X_H^{-1})] \cdot (\psi - X\beta)' H^{-1} (\psi - X\beta) \cdot H^{-1} \\
& - \frac{2}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+1+\alpha \text{tr}(X_H^{-1})] \cdot (\psi - X\beta)' H^{-1} (\psi - X\beta) \cdot H^{-1} X_H^{-1} \\
& - \frac{1}{n} \left( \frac{1}{n} + \frac{1}{d} \right) [k+2+\alpha \text{tr}(X_H^{-1})] [H^{-1} X_H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} + H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} X_H^{-1} X_H^{-1}] \\
& - \frac{2}{n} \left( \frac{1}{n} + \frac{1}{d} \right) \cdot (\psi - X\beta)' H^{-1} X_H^{-1} (\psi - X\beta) \cdot H^{-1} \\
& - 2 \left( \frac{1}{n} + \frac{1}{d} \right) \sigma^2 \cdot (\psi - X\beta)' H^{-1} X_H^{-1} (\psi - X\beta) \cdot H^{-1} X_H^{-1} \\
& - \left( \frac{1}{n} + \frac{1}{d} \right) \sigma^2 [H^{-1} X_H^{-1} X_H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} + H^{-1} (\psi - X\beta) (\psi - X\beta)' H^{-1} X_H^{-1} X_H^{-1}]
\end{aligned}
\tag{4.21}$$

where  $\sigma_w^2 = (\sigma_e^2 - 2\psi'\beta + \beta'X\beta)$ . Expression 4.21 gives, up to terms of order in probability  $(1/n^2)$ , the difference in mean square errors of the estimators  $\tilde{\beta}$  and  $\hat{\beta}$ . By inspection of 4.21, it can be seen that in general  $\tilde{\beta}$  does not have necessarily smaller mean square error than  $\hat{\beta}$ .

However, if the matrices  $\Sigma$  and  $H^{-1}$  commute, the traces of all matrices entering in 4.21 are positive. Therefore, if  $\Sigma$  and  $H^{-1}$  commute, then  $\tilde{\beta}$  has smaller mean square error than  $\hat{\beta}$ . We state this result in the following theorem.

Theorem 1 Consider the following estimators of  $\beta$  in Model 4.1:

$$\begin{aligned}\hat{\beta} &= \hat{H}^{-1}\hat{N} \\ \tilde{\beta} &= (\hat{H} + \hat{G})^{-1}(\hat{N} + \hat{D}) .\end{aligned}$$

If  $\Sigma$ , the covariance matrix of the errors of  $x$ , and  $H^{-1}$ , the matrix related to  $\hat{H}$  by 4.3 commute, that is, if

$$\Sigma H^{-1} = H^{-1} \Sigma$$

then  $\tilde{\beta}$  has smaller mean square error than  $\hat{\beta}$ . The difference in mean square errors up to terms of order in probability  $(1/n^2)$  is given by 4.21.

Observe that the above condition for  $\tilde{\beta}$  to have smaller mean square error than  $\hat{\beta}$  is sufficient but not necessary. Therefore, even in cases when  $\Sigma$  and  $H^{-1}$  do not commute it is possible for  $\tilde{\beta}$  to have smaller mean square error than  $\hat{\beta}$ . Observe also that if both  $H^{-1}$  and  $\Sigma$  are scalars, then the condition of commutativity required by Theorem 1 holds. In this particular case  $\tilde{\beta}$  has always smaller mean square error than  $\hat{\beta}$ . We obtain, therefore, the following Corollary to Theorem 1.

Corollary 1 Let  $k = 1$  in Model 1 so that  $\Sigma$  and  $H^{-1}$  are scalars. Then, under these conditions  $\Sigma H^{-1} = H^{-1} \Sigma$ , which by Theorem 1 is sufficient for  $\tilde{\beta}$  to have smaller mean square error than  $\hat{\beta}$ . If in this case we write  $\Sigma = \sigma_u^2$ ,  $\psi = \sigma_{ue}$  and  $H = S_{XX}$ , then

$$\begin{aligned}
E(\tilde{\beta}-\beta)^2 - E(\hat{\beta}-\beta)^2 = & -\frac{4}{ndS_{XX}} \left[ 1 + \alpha \frac{\sigma_u^2}{S_{XX}} \right] [\sigma_w^2 \sigma_u^2 + \sigma_{uw}^2] \\
& - \frac{8\alpha^2 \sigma_u^2 \sigma_{uw}^2}{n^2 S_{XX}^3} \left[ 1 + \frac{\sigma_u^2}{S_{XX}} \right] \\
& - \frac{12\sigma_{uw}^2}{n^2 S_{XX}^2} \left[ 1 + 2\alpha \frac{\sigma_u^2}{S_{XX}} + \frac{\alpha^2 (\sigma_u^2)^2}{S_{XX}^2} \right] \\
& - \frac{4\sigma_u^2}{n^2 S_{XX}^2} \left[ 2 + \alpha \frac{\sigma_u^2}{S_{XX}} \right] \left[ \sigma_w^2 + \frac{\alpha \sigma_u^2 \sigma_w^2}{S_{XX}} + \frac{\alpha \sigma_{uw}^2}{S_{XX}} \right]
\end{aligned} \tag{4.22}$$

where  $\sigma_w^2 = \sigma_e^2 - 2\beta\sigma_{ue} + \beta^2\sigma_u^2$ ,  $\sigma_{uw}^2 = (\sigma_{ue} - \beta\sigma_u^2)^2$  and  $\alpha = (1 + n/d)$ .

#### B. Estimators with Finite Expectations

The estimators in the previous section do not necessarily have finite expectations. However, it will be shown that simple modifications to the estimators already obtained will produce estimators with finite moments. The asymptotic properties of the estimators so modified are equivalent to those of the estimators previously studied.

We shall introduce some notation to be used throughout the section.

Let  $A$  be a  $p \times p$  positive definite symmetric matrix of random elements, and  $B$  a  $p \times p$  positive semidefinite symmetric matrix of random elements.

Let  $\lambda =$  maximum characteristic root of  $A$ ,  
 $\lambda^* =$  minimum characteristic root of  $A$ ,  
 $\mu =$  maximum characteristic root of  $(A + B)$

and  $\mu^* =$  minimum characteristic root of  $(A + B)$ .

We shall prove the following two theorems:

Theorem 2  $E\{|1/\lambda^*|^r\} < +\infty$  implies the existence of the  $r$ th moments of the elements of  $(A + B)^{-1}$ .

Theorem 3  $E\{|1/\lambda^*|^{2r}\} < +\infty$  and  $E\{|\lambda|^r\} < +\infty$ ,  $r > 0$ , imply existence of the  $r$ th moments of the elements of  $C^{-1}AC^{-1}$ , where  $C = A + B$ .

To prove Theorems 2 and 3 we show some preliminary results.

Lemma 1 Let  $a_{ij}$  be the element on the  $i$ th row,  $j$ th column of  $A$ . Then

$$\lambda \geq |a_{ij}|, \quad 1 \leq i, j \leq p.$$

Proof If  $A$  is a positive definite symmetric matrix then there exists an orthogonal matrix  $Q$  such that,

$$\{Q A Q'\}_{ij} = \sum_{k=1}^p \sum_{\ell=1}^p q_{i\ell} a_{\ell k} q_{kj}^* = \{D\}_{ij} = d_{ij} \quad (4.23)$$

where  $\{Q'\}_{ij} = q_{ij}^*$ , and  $D$  is a diagonal matrix whose elements are the characteristic roots of  $A$ . From 4.23

$$a_{ij} = \{Q'DQ\}_{ij} = \sum_{k=1}^p \sum_{\ell=1}^p q_{i\ell}^* d_{\ell k} q_{kj} \quad (4.24)$$

with  $q_{i\ell}^* = q_{\ell i}$ , and  $d_{\ell k} = 0$  for  $\ell \neq k$ . Now,  $\lambda \geq d_{kk}$ ,  $1 \leq k \leq p$ .

Therefore

$$a_{ij} = \sum_{k=1}^p q_{ik}^* d_{kk} q_{kj} = \sum_{k=1}^p q_{ki} d_{kk} q_{kj} \quad (4.25)$$

$$a_{ii} = \sum_{k=1}^p q_{ki} d_{kk} q_{ki} \leq \left( \sum_{k=1}^p q_{ki}^2 \right) \lambda = \lambda, \quad 1 \leq i \leq p. \quad (4.26)$$

since  $\sum_{k=1}^p q_{ki}^2 = 1$ . Similarly

$$\begin{aligned}
 |a_{ij}| &= \left| \sum_{k=1}^p q_{ki} d_{kk} q_{kj} \right| \leq \sum_{k=1}^p |q_{ki} q_{kj} d_{kk}| \\
 &\leq \sum_{k=1}^p |q_{ki} q_{kj}| d_{kk} \\
 &\leq \sum_{k=1}^p |q_{ki} q_{kj}| \lambda \\
 &\leq |\cos \theta| \lambda \\
 &\leq \lambda
 \end{aligned} \tag{4.27}$$

where  $\theta$  is the angle between the vectors of absolute values of  $q_i$  and  $q_j$ , the  $i$ th and  $j$ th column vectors of  $Q$ , respectively. i.e.,  $q_i = (|q_{1i}|, |q_{2i}|, \dots, |q_{pi}|)'$ ,  $i = 1, 2, \dots, p$ .

Lemma 2  $C = A + B$  is a positive definite symmetric matrix such that  $\mu \geq \lambda$  and  $\mu^* \geq \lambda^*$ .

Proof To prove Lemma 2 we use the following theorem given on page 65 of Bodewig (1956). In terms of our notation:

Theorem (Bodewig (1956)) The symmetric form  $A$  has  $p$  real eigenvalues  $\lambda_i$  and real eigenvectors  $x_i$  which are characterized by the following properties.

1) The largest eigenvalue  $\lambda_p$  is the maximum of the quadratic form  $x'Ax$  for all vectors of the length 1, that is

$$\lambda_p = \max \frac{x'Ax}{x'x}.$$

The vector  $x$  which gives the maximum value is the eigenvector corresponding to  $\lambda_p$  :

$$\lambda_p = (x_p' A x_p) / (x_p' x_p) \quad \text{and} \quad A x_p = \lambda_p x_p .$$

2) The eigenvalue  $\lambda_i$  is the maximum of the form  $x' A x$  for all vectors  $x$  of the length 1 for which

$$x' x_1 = x' x_2 = x' x_3 = \dots = x' x_{i-1} = 0 .$$

The vector  $x_i$  for which the form assumes the maximum is the eigenvector belonging to  $\lambda_i$  .

3) The smallest eigenvalue  $\lambda_1$  is the minimum of the quadratic form  $x' A x$  for all vectors of length 1, that is

$$\lambda_1 = \min \frac{x' A x}{x' x} .$$

We proceed now to prove Lemma 2. By definition

A positive definite  $\Rightarrow x' A x > 0$  for all vectors  $x$  ,

B positive semidefinite  $\Rightarrow x' B x \geq 0$  for all vectors  $x$  .

Therefore

$$x' (A+B) x \geq x' A x > 0 \quad \text{for all vectors } x ,$$

so that,  $A + B$  is a positive definite symmetric matrix. Now, let

$$R = \{x : x' x = 1\} .$$

i.e.,  $R$  is the set of all  $p$ -dimensional vectors of length 1. Assume that



for a vector  $x_0 \in R$

$$\frac{x_0'Ax_0}{x_0'x_0} = \lambda = \max_{x \in R} \frac{x'Ax}{x'x} . \quad (4.28)$$

By the preceding theorem,  $\lambda$  is the maximum characteristic root of  $A$ .

However,

$$\frac{x_0'(A+B)x_0}{x_0'x_0} = \frac{x_0'Ax_0}{x_0'x_0} + \frac{x_0'Bx_0}{x_0'x_0} \quad (4.29)$$

is not necessarily the maximum of

$$\frac{x'(A+B)x}{x'x} \quad (4.30)$$

If  $x_1 \in R$ , is such that 4.30 is a maximum, we deduce from 4.29

$$\mu = \frac{x_1'(A+B)x_1}{x_1'x_1} \geq \max_{x \in R} \frac{x'(A+B)x}{x'x} \geq \frac{x_0'(A+B)x_0}{x_0'x_0} \geq \lambda . \quad (4.31)$$

We now prove that  $\mu^* \geq \lambda^*$ . Assume that for a vector  $x_0^* \in R$

$$\frac{x_0^{*'}Ax_0^*}{x_0^{*'}x_0^*} = \lambda^* = \min_{x \in R} \frac{x'Ax}{x'x} . \quad (4.32)$$

By the preceding theorem,  $\lambda^*$  is the minimum characteristic root of  $A$ .

However

$$\frac{x_0^{*'}(A+B)x_0^*}{x_0^{*'}x_0^*} = \frac{x_0^{*'}Ax_0^*}{x_0^{*'}x_0^*} + \frac{x_0^{*'}Bx_0^*}{x_0^{*'}x_0^*} \quad (4.33)$$

is not necessarily the minimum of

$$\frac{x'(A+B)x}{x'x} . \quad (4.34)$$

If  $x_1^* \in R$ , is such that 4.33 is a minimum, we deduce from 4.32

$$\lambda^* = \frac{x_0^{*'} A x_0^*}{x_0^{*'} x_0^*} = \min_{x \in R} \frac{x' A x}{x' x} \leq \frac{x_1^{*'} (A+B) x_1^*}{x_1^{*'} x_1^*} = \mu^* . \quad (4.35)$$

Equations 4.31 and 4.35 prove Lemma 2.

Lemma 3 Let  $\mu_1, \mu_2, \dots, \mu_p$  be the characteristic roots of  $C = A + B$ . Then  $C^{-1}$  has characteristic roots  $1/\mu_1, 1/\mu_2, \dots, 1/\mu_p$ . Therefore,  $1/\mu$  is the minimum characteristic root of  $C^{-1}$  and  $1/\mu^*$  its maximum characteristic root.

Proof  $C$  positive definite symmetric implies the existence of an orthogonal matrix  $Q$  such that

$$Q C Q' = D \quad (4.36)$$

where  $D$  is a nonsingular diagonal matrix with elements the characteristic roots of  $C$ . Therefore, the following relation holds:

$$D^{-1} = Q'^{-1} C^{-1} Q^{-1} .$$

But  $Q^{-1} = Q'$  and  $Q'^{-1} = Q$  since  $Q$  is orthogonal, so that

$$D^{-1} = Q C^{-1} Q' .$$

Since  $D^{-1}$  is diagonal with positive elements and  $Q$  is orthogonal, the elements of  $D^{-1}$  are the characteristic roots of  $C^{-1}$ . Also, if  $\mu_1, \dots, \mu_p$  are the elements of  $D$ ,  $1/\mu_1, \dots, 1/\mu_p$  are the elements of

$D^{-1}$ . It also follows that if  $\mu = \max\{\mu_1, \dots, \mu_p\}$ , then  $1/\mu = \min\{1/\mu_1, \dots, 1/\mu_p\}$  with corresponding expressions for  $\mu^*$ .

Lemma 4 Let  $C = A + B$  and let  $c^{ij}$  denote the element on the  $i$ th row,  $j$ th column of  $C^{-1}$ . Then, for any possible event

$$|c^{ij}| \leq 1/\mu^* \leq 1/\lambda^*, \quad 1 \leq i, j \leq p.$$

Proof By Lemma 3  $1/\mu^*$  is the maximum characteristic root of  $C^{-1}$ . Therefore, by Lemma 1

$$|c^{ij}| \leq 1/\mu^*, \quad 1 \leq i, j \leq p.$$

Also, because of Lemma 2  $\mu^* \geq \lambda^*$ . Then the desired result follows.

Using Lemmas 1 through 4 we shall prove Theorem 2.

Theorem 2  $E\{|1/\lambda^*|^r\} < +\infty$  implies the existence of the  $r$ th moments of the elements of  $(A+B)^{-1}$ .

Proof We have to show that

$$E\{|1/\lambda^*|^r\} < +\infty \tag{4.37}$$

implies the existence of the moments of  $c^{ij}$ ,  $1 \leq i, j \leq p$ , the element on the  $i$ th row,  $j$ th column of  $C^{-1}$ , where  $C = A + B$ . By Lemma 4, it is always true that

$$|c^{ij}| \leq 1/\mu^* \leq 1/\lambda^*.$$

Therefore

$$E\{|c^{ij}|^r\} \leq E\{|1/\lambda^*|^r\} < +\infty, \quad r > 0.$$

We shall now prove Theorem 3.

Theorem 3  $E\{|1/\lambda^*|^{2r}\} < +\infty$  and  $E\{|\lambda|^r\} < +\infty$  imply existence of the  $r$ th moments of the elements of  $C^{-1}AC^{-1}$ .

Proof Let  $w_{ij}$  be the element on the  $i$ th row,  $j$ th column of  $C^{-1}AC^{-1}$ . Then

$$w_{ij} = \sum_{\ell=1}^p \sum_{k=1}^p c^{i\ell} a_{\ell k} c^{kj}$$

$$|w_{ij}| \leq \sum_{\ell=1}^p \sum_{k=1}^p |c^{i\ell}| |a_{\ell k}| |c^{kj}|$$

$$\leq \sum_{\ell=1}^p \sum_{k=1}^p \frac{1}{\mu^*} \lambda \frac{1}{\mu^*}$$

$$\leq p^2 \left(\frac{1}{\mu^*}\right)^2 \lambda$$

$$\leq p^2 \left(\frac{1}{\lambda^*}\right)^2 \lambda \quad .$$

The theorem follows from the last inequality.

Let us illustrate the use of Theorems 2 and 3.

Let  $G$  be a  $k \times k$  matrix of random variables such that all of its elements are  $O_p(1)$ . Assume that the  $r$ th moments of the elements of  $G^{-1}$  do not exist. Let  $P$  be a  $k \times k$  positive definite symmetric matrix of random variables all of whose elements are  $O_p(1/n)$ . Then

$$\begin{aligned} G^{-1} &= (G^{-1}G', -1)G' = (G'G)^{-1}G' \\ &= [(G'G+P) - P]^{-1}G' \\ &= \{(G'G+P)[I - (G'G+P)^{-1}P]\}^{-1}G' \end{aligned}$$

$$\begin{aligned}
&= [I - (G'G+P)^{-1}P]^{-1}(G'G+P)^{-1}G' \\
&= [I + (G'G+P)^{-1}P + O_p(1/n^2)](G'G+P)^{-1}G'
\end{aligned}$$

$$G^{-1} = (G'G+P)^{-1}G' + (G'G+P)^{-1}P(G'G+P)^{-1}G' + O_p(1/n^2) \quad (4.38)$$

We observe that  $G'G$  is a positive semidefinite symmetric matrix in expression 4.38. Let

$$D = (G'G+P)^{-1}G' + (G'G+P)^{-1}P(G'G+P)^{-1}G' \quad (4.39)$$

then  $G$  and  $D$  are equivalent except for terms of  $O_p(1/n^2)$ . If  $G$  is an estimator whose elements do not have  $r$ th finite moments, we can replace it by  $D$ .  $D$  has to  $O_p(1/n)$  the same asymptotic properties of  $G$ . In addition, by choosing adequately the matrix  $P$ , its elements will have finite  $r$ th moments.

Example 1 Let  $S$  be a  $p \times p$  random matrix which has a Wishart distribution  $W_p(\Sigma, d = n-1)$ . James (1954), has shown that the joint density of the characteristic roots of  $S$ ,  $\lambda_1, \dots, \lambda_p$ , can be written as follows:

$$f(\lambda_1, \dots, \lambda_p) = k(d, p) \left( \prod_{i=1}^p \lambda_i \right)^{\frac{1}{2}(d-p-1)} h(\lambda_1, \dots, \lambda_p) \text{ on } R^p$$

where  $k(d, p)$  is a constant depending upon  $d$  and  $p$ . Then

$$\int \dots \int \left( \prod_{i=1}^p \lambda_i \right)^{\frac{1}{2}(d-p-1)} h(\lambda_1, \dots, \lambda_p) d\lambda_1 \dots d\lambda_p = \frac{1}{k(d, p)} \quad (4.40)$$

Now

$$E \left( \prod_{i=1}^p \lambda_i^{-r} \right) = k(d, p) \int \dots \int \left( \prod_{i=1}^p \lambda_i \right)^{\frac{1}{2}(d-2r-p-1)} h(\lambda_1, \dots, \lambda_p) d\lambda_1 \dots d\lambda_p$$

$$= \frac{k(d,p)}{k(d-4r,p)} , \quad (4.41)$$

where the integral in 4.41 exists if  $n \geq 2r + p + 2$ . The existence of the integral in 4.41 implies the existence of

$$E\{(1/\lambda_i)^r\} , \quad 1 \leq i \leq p .$$

It is also easily seen that  $E\{\lambda_i^r\}$ ,  $1 \leq i \leq p$  exists for all  $r = 1, 2, \dots$ . Therefore, if we write

$$P = S/n ,$$

the elements of  $P$  will be  $O_p(1/n)$ . Letting

$$A = P ,$$

and

$$B = G'G ,$$

all the conditions to apply Theorems 2 and 3 are satisfied. The elements of  $D$  will have finite  $r$ th moments provided the  $r$ th moments of the elements of  $G$  exist and  $n \geq 2r + p + 2$ .

### C. Estimation in a Simultaneous Equations Problem

#### 1. Introduction

Let us assume a linear model containing  $G$  structural relations. The  $i$ th relation at time  $t$  may be written

$$\beta_{i1}y_{1t} + \dots + \beta_{iG}y_{Gt} + \gamma_{i1}x_{1t} + \dots + \gamma_{iK}x_{Kt} = u_{it} ,$$

$$i = 1, 2, \dots, G , \quad t = 1, 2, \dots, n$$

where the  $y_{it}$  denote endogenous variables and the  $x_{it}$  predetermined variables. The system of  $G$  equations above can be written in matrix form as

$$By_t + \Gamma x_t = u_t, \quad t = 1, 2, \dots, n,$$

where  $B$  is the  $G \times G$  matrix of coefficients of the endogenous variables (which we assume to be nonsingular),  $\Gamma$  the  $G \times K$  matrix of coefficients of the predetermined variables,  $y_t$ ,  $x_t$ ,  $u_t$  are column vectors of  $G$ ,  $K$  and  $G$  elements respectively. Assuming that the  $B$  matrix is nonsingular, the reduced form of the model may be written

$$y_t = \pi x_t + v_t$$

where  $\pi$  is a  $G \times K$  matrix of reduced form coefficients and  $v_t$  a column vector of  $G$  reduced form disturbances

$$\pi = -B^{-1}\Gamma,$$

$$v_t = B^{-1}u_t.$$

If an equation of the system is to be identified, we must have zero-nonzero restrictions. Assume that the restrictions on the coefficients of the first relation are such that  $G_1$  is the number of endogenous variables and  $K_1$  the number of predetermined variables which appear in the relation with non-zero coefficients. Then  $G_2 = G - G_1$  and  $K_2 = K - K_1$  are the numbers of variables per class not included in the first relation. It can be shown that the relation in question is identified if  $K_2 \geq G_1 - 1$ . Or in other words, the condition implies that the number of predetermined variables excluded from the relation must be at least as great as the number of endogenous variables less one.

Using the "a priori" restrictions on the coefficients of the first relation, we can write it as

$$\beta_{11}y_{1t} + \dots + \beta_{1G_1}y_{G_1t} + \gamma_{11}x_{1t} + \dots + \gamma_{1K_1}x_{K_1t} = u_{1t},$$

$$t = 1, 2, \dots, n.$$

Assume the above relation is identified. Normalizing by setting  $\beta_{11} = 1$ , we can write

$$y_{1t} = -\beta_{12}y_{2t} - \dots - \beta_{1G_1}y_{G_1t} - \gamma_{11}x_{1t} - \dots - \gamma_{1K_1}x_{K_1t} + u_{1t}$$

$$t = 1, 2, \dots, n,$$

or in matrix form

$$y_1 = -Y_2\beta'_2 - X_1\gamma'_1 + u_1 \quad (4.42)$$

where

$$y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \end{bmatrix}, \quad Y_2 = \begin{bmatrix} y_{21} & \dots & y_{G_1 1} \\ \vdots & & \vdots \\ y_{2n} & \dots & y_{G_1 n} \end{bmatrix}, \quad \beta'_2 = \begin{bmatrix} \beta_{12} \\ \vdots \\ \beta_{1G_1} \end{bmatrix},$$

$$X_1 = \begin{bmatrix} x_{11} & \dots & x_{K_1 1} \\ \vdots & & \vdots \\ c_{1n} & \dots & x_{K_1 n} \end{bmatrix}, \quad \gamma'_1 = \begin{bmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1K_1} \end{bmatrix}.$$

One of the methods of estimation of the parameters of the structural relation 4.42, namely: two-stage least squares, is based on the following



idea. Replace the matrix  $Y_2$  of 4.42 by an estimated matrix  $\hat{Y}_2 = X(X'X)^{-1}X'Y_2$ , where

$$X = (X_1, X_2) ,$$

$X_2$  being the matrix of predetermined variables not included in the equation.

We can also write  $Y_2 = \hat{Y}_2 + \hat{V}_2$ , where  $\hat{V}_2$  denotes the matrix of estimated reduced form residuals for the  $G_1 - 1$  endogenous variables appearing on the right hand side of 4.42. Expression 4.42 may be rewritten

$$y_1 = -\hat{Y}_2\beta_2' - X_1\gamma_1' + (u_1 - \hat{V}_2\beta_2') . \quad (4.43)$$

Applying least squares to this relation we obtain

$$\begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_1' \end{bmatrix} = - \begin{bmatrix} \hat{Y}_2'\hat{Y}_2 & \hat{Y}_2'X_1 \\ X_1'Y_2 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2'y_1 \\ X_1'y_1 \end{bmatrix} .$$

We shall show that two-stage least squares described above, can be considered as an errors in variables problem.

## 2. Connection with the errors in variables approach

Let us consider relation 4.43 above and let the  $X$ 's be fixed, then

$$y_1 = -\hat{Y}_2\beta_2' - X_1\gamma_1' + e_1 ,$$

where  $e_1 = u_1 - \hat{V}_2\beta_2'$  and  $\hat{Y}_2 = X(X'X)^{-1}X'Y_2$ . Let us examine the covariance structure between column vectors of the matrix  $(y_1, \hat{Y}_2)$ .

In the remaining of the present discussion we shall write  $y_t = (y_{1t}, y_{2t})$ ,  $t = 1, \dots, n$ . i.e.,  $y_t$  is the  $t$ th row of the matrix  $(y_1, Y_2) = Y$ , say. We recall that by virtue of the reduced form

$$\begin{aligned}
E(y_t - Ey_t)'(y_s - Ey_s) &= E(y_t - \pi x_t)'(y_s - \pi x_s) \\
&= E(v_t' v_s) = \delta_{ts} \phi, \quad 1 \leq t, s \leq n,
\end{aligned} \tag{4.44}$$

where  $\delta_{ts} = 0$  or 1 when  $t \neq s$  or  $t = s$ , respectively, and the matrix  $\phi$  is of order  $G_1 \times G_1$ . Then the matrix of errors in 4.43 is given by

$$(y_1 - Ey_1, \hat{Y}_2 - E\hat{Y}_2) \tag{4.45}$$

Now, letting  $X(X'X)^{-1}X' = M$ , we have

$$\hat{Y}_2 - E\hat{Y}_2 = X(X'X)^{-1}X'(Y_2 - EY_2) = MV_2.$$

Then expression 4.45 can be written

$$(v_1, MV_2),$$

where  $v_1$  is a  $n \times 1$  vector and  $V_2$  a  $n \times (G_1 - 1)$  matrix. Let  $V_{2j}$  be the  $j$ th column of  $V_2$ ,  $j = 2, \dots, G_1$ . Then, it is easily seen that

$$E\{v_1 v_{2j}' M\} = M\phi_{1j}, \quad j = 2, \dots, G_1, \tag{4.46}$$

and

$$E\{MV_{2j} V_{2s}' M\} = M\phi_{js}, \quad 2 \leq j, s \leq G_1, \tag{4.47}$$

where in 4.46 and 4.47  $\phi_{js}$  is the element on the  $j$ th row,  $s$ th column of the matrix  $\phi$ .

We see from the results in 4.46 and 4.47 that the errors given by 4.45 are not necessarily time uncorrelated. This would be the case if  $M$  was the  $n \times n$  identity matrix. Therefore, under the above conditions, the errors in variables approach cannot be used. However, if we are able to

transform the original problem in such a way that the new errors are now time uncorrelated, with constant covariance structure in time, we would be able to apply the errors in variables approach.

We shall prove the following theorem:

Theorem 4      The two-stage least squares estimators remain invariant under any orthogonal transformation of model 4.42. That is, if  $Q$  is any  $n \times n$  orthogonal matrix the two-stage least squares estimators obtained under the model

$$y_1 = -Y_2\beta_2' - X_1\gamma_1' + u_1$$

are identical in value with the two-stage least squares estimators obtained with the model

$$z_1 = -Z_2\beta_2' - W_1\gamma_1' + w_1 ,$$

where  $z_1 = Qy_1$  ,  $Z_2 = QY_2$  ,  $W = QX = (W_1, W_2) = (QX_1, QX_2)$  and  $w_1 = Qu_1$  .

Proof      First of all notice that

$$\begin{aligned}\hat{Z}_2 &= W(W'W)^{-1}W'Z_2 \\ &= QX(X'Q'QX)^{-1}X'Q'QY_2 \\ &= QX(X'X)^{-1}X'Y_2 \\ &= Q\hat{Y}_2 .\end{aligned}$$

Therefore, the two-stage least squares estimators are obtained by solving the system

$$\begin{aligned}
\begin{bmatrix} \hat{\beta}_2' \\ \gamma_1' \end{bmatrix} &= - \begin{bmatrix} \hat{Z}_2' \hat{Z}_2 & \hat{Z}_2' W_1 \\ W_1' \hat{Z}_2 & W_1' W_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Z}_2' z_1 \\ W_1' z_1 \end{bmatrix} \\
&= - \begin{bmatrix} \hat{Y}_2' Q' Q \hat{Y}_2 & \hat{Y}_2' Q' Q X_1 \\ X_1' Q' Q \hat{Y}_2 & X_1' Q' Q X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2' Q' Q y_1 \\ X_1' Q' Q y_1 \end{bmatrix} \\
&= - \begin{bmatrix} \hat{Y}_2' \hat{Y}_2 & \hat{Y}_2' X_1 \\ X_1' \hat{Y}_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_2' y_1 \\ X_1' y_1 \end{bmatrix}
\end{aligned}$$

which proves the theorem.

Now, observe that  $M = X(X'X)^{-1}X'$  is an  $n \times n$  symmetric idempotent matrix; then its characteristic roots are, all of them, either 0 or 1. Therefore, from matrix theory we know that there exists a  $n \times n$  orthogonal matrix diagonalizing  $M$ . Let  $Q$  be such a matrix, then

$$QMQ' = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} I_{K \times K} & 0 \\ 0 & 0 \end{bmatrix}. \quad (4.48)$$

If in the proof of Theorem 4 we choose  $Q$  satisfying 4.48, then the transformed model

$$z_1 = -Z_2 \beta_2' - W_1 \gamma_1' + u_1$$

is such that the matrix of errors

$$(z_1 - Ez_1, \hat{z}_2 - E\hat{z}_2)$$

has uncorrelated rows. We can prove this assertion by noting that this matrix of errors can be written in terms of the original variables as follows:

$$(Qv_1, QV_2),$$

where  $v_1 = y_1 - Ey_1$  and  $V_2 = Y_2 - EY_2$ . Then

$$\begin{aligned} E(Qv_1 V_2' MQ') &= QMQ' \phi_{1j} = \begin{bmatrix} I_{K \times K} & 0 \\ 0 & 0 \end{bmatrix} \cdot \phi_{1j}, \\ E(QV_j V_s' MQ') &= QMQ' \phi_{js} = \begin{bmatrix} I_{K \times K} & 0 \\ 0 & 0 \end{bmatrix} \phi_{js}, \end{aligned} \quad (4.49)$$

$2 < j, s \leq G_1,$

where we have used the results given by 4.46 and 4.47. Therefore, if  $Q$  satisfies

$$QMQ' = \begin{bmatrix} I_{K \times K} & 0 \\ 0 & 0 \end{bmatrix},$$

this implies that under the transformation the new errors are uncorrelated in time. Moreover, because  $Q$  is orthogonal

$$\begin{aligned} E(y_{1t} - Ey_{1t}, Y_{2t} - EY_{2t})' (y_{1t} - Ey_{1t}, Y_{2t} - EY_{2t}) \\ = E(z_{1t} - Ez_{1t}, \hat{z}_{2t} - E\hat{z}_{2t})' (z_{1t} - Ez_{1t}, \hat{z}_{2t} - E\hat{z}_{2t}) \\ = \phi, \quad t = 1, 2, \dots, K. \end{aligned} \quad (4.50)$$

Also from the discussion given below, it will be clear that the remaining

$n - K$  relations are error identities so that they do not give any additional information about the parameters. Therefore, they can be eliminated for purposes of estimation.

Let us partition  $Q$  as follows:

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$$

where  $Q_1$  is a  $K \times n$  matrix and  $Q_2$  is  $(n-K) \times n$ . Now

$$QM'Q' = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} M(Q_1', Q_2') = \begin{bmatrix} Q_1 M Q_1' & Q_1 M Q_2' \\ Q_2 M Q_1' & Q_2 M Q_2' \end{bmatrix} = \begin{bmatrix} I_{K \times K} & 0 \\ 0 & 0 \end{bmatrix}. \quad (4.51)$$

Corresponding to the above partition, we obtain a similar partition of  $y_1$

$$Q y_1 = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} y_1 = - \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} Y_2 \beta_2' - \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} X_1 \gamma_1' + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} u_1$$

or

$$\begin{aligned} Q_1 y_1 &= - Q_1 Y_2 \beta_2' - Q_1 X_1 \gamma_1' + Q_1 u_1, \\ Q_2 y_1 &= - Q_2 Y_2 \beta_2' - Q_2 X_1 \gamma_1' + Q_2 u_1. \end{aligned}$$

Let  $Q_1 y_1 = z_{11}$ ,  $Q_2 y_1 = z_{12}$ ,  $Q_1 Y_2 = Z_{21}$ , etc., the last two expressions can be written

$$\begin{aligned} z_{11} &= - Z_{21} \beta_2' - W_1 \gamma_1' + w_{11}, \\ z_{12} &= - Z_{22} \beta_2' - W_2 \gamma_1' + w_{12}. \end{aligned} \quad (4.52)$$

We have several observations with respect to the above results:

(i) if 4.51 holds  $Q_1$  is an orthogonal basis for the column space of the  $n \times K$  matrix  $X$ ;

(ii) as a consequence of (i) the matrices  $\hat{Q}_2 \hat{Y}_2$  and  $\hat{Q}_2 X_1$  are  $K \times G_1$  and  $K \times K_1$  zero matrices, respectively;

(iii) the two stage least squares estimating equation given in terms of the transformed variables reduces to

$$z_{11} = - \hat{Z}_{21} \beta_2' - W_1 \gamma_1' + w_{11} . \quad (4.53)$$

(iv) by virtue of (ii)

$$z_{12} = - \hat{Z}_{22} \beta_2' - W_2 \gamma_1' + w_{12}$$

is a  $(n-K) \times 1$  vector of error identities. They do not give any further information about the parameters to be estimated;

(v) if the equation is identified we must have  $K \geq G_1 + K_1 - 1$  ;

(vi) finally, because of the way  $\hat{Q}$  was chosen, the errors of  $\hat{Z}_{21}$  have the same covariance structure as the reduced form errors  $\hat{Y}_2$  and the errors of  $\hat{Z}_{21}$  are time uncorrelated.

Therefore, we have shown the following theorem.

Theorem 5      There exists an orthogonal transformation which transforms model 4.42 into a new model whose two stage least squares estimating equations 4.53, determine a standard errors in variables model. The covariance structure of the errors of the variables in 4.53 is identical with the covariance structure of the errors of the reduced form. The new system is a system of  $K$  equations in  $G_1 + K_1$  parameters.

### 3. The estimation method

As a first step to obtain nearly unbiased estimators of the parameters of Model 4.42 we shall prove Theorem 6.

Theorem 6 An unbiased estimator of  $\phi$ , the covariance structure of the reduced form residuals is given by

$$\hat{\phi} = \frac{Y'[I - X(X'X)^{-1}X']Y}{n-K} \quad (4.54)$$

where  $Y$  is the  $n \times G$  matrix of endogenous variables in the equation to be estimated.

Proof: In terms of the reduced form parameters,  $Y = X\pi + V$  where  $V = (v_1, v_2)$  is the  $n \times G_1$  matrix of reduced form residuals. Let  $M = X(X'X)^{-1}X'$ , then

$$\begin{aligned} E\{(n-K)\hat{\phi}\} &= E\{Y'(I-M)Y\} \\ &= E\{(X\pi+V)'(I-M)(X\pi+V)\} \\ &= E\{\pi'X'(I-M)X\pi\} + E\{\pi'X'(I-M)V\} + E\{V'(I-M)X\pi\} + E\{V'(I-M)V\} \\ &= E\{V'(I-M)V\} . \end{aligned}$$

Now, let  $v_{ij}$  and  $m_{ij}$  be the elements on the  $i$ th row,  $j$ th column of the matrices  $V$  and  $(I-M)$ , respectively. Consider the expected value of the  $(qs)$  element of  $V'(I-M)V$ ; i.e.,

$$\begin{aligned} E\left[\sum_{i=1}^n \sum_{j=1}^n v_{iq} v_{js} m_{ij}\right] &= \sum_{i=1}^n \sum_{j=1}^n E(v_{iq} v_{js}) \cdot m_{ij} \\ &= \sum_{i=1}^n \sum_{j=1}^n \delta_{ij} \phi_{qs} m_{ij} \\ &= \phi_{qs} \sum_{i=1}^n m_{ii} \\ &= \phi_{qs} \{\text{rank of } (I-M)\} \\ &= (n-K)\phi_{qs} . \end{aligned}$$



Therefore,  $E\{(n-K)\hat{\phi}\} = (n-K)\phi$  and the conclusion of the theorem follows.

Based on Theorem 6 consider the expression

$$\hat{H} = \begin{bmatrix} \frac{Y_2' M Y_2}{n} & \frac{Y_2' M X_1}{n} \\ \frac{X_1' M Y_2}{n} & \frac{X_1' X_1}{n} \end{bmatrix} - K \begin{bmatrix} \frac{Y_2' (I-M) Y_2}{n(n-K)} & 0 \\ 0 & 0 \end{bmatrix} ; \quad (4.55)$$

we shall prove the following result.

Theorem 7 Let  $(X'X/n) = \Sigma_{XX}$ ,  $(X'X_1/n) = \Sigma_{XX_1}$  and  $(X_1'X_1/n) = \Sigma_{X_1X_1}$ . Then the expected value of  $\hat{H}$  given by 4.55 is

$$\begin{bmatrix} \pi_2' \Sigma_{XX} \pi_2 & \pi_2' \Sigma_{XX_1} \\ \Sigma_{XX_1}' \pi_2 & \Sigma_{X_1X_1} \end{bmatrix} = H , \quad (4.56)$$

where  $\pi_2$  is a part of the reduced form matrix of coefficients corresponding to the endogenous variables in  $Y_2$ .

Proof: We first observe that  $\frac{1}{n-K} [Y_2'(I-M)Y_2]$  can be written as follows

$$\frac{Y_2'(I-M)Y_2}{n-K} \frac{(n-K)+K}{n} = \frac{Y_2'(I-M)Y_2}{(n-K)n} (n-K) + \frac{K}{n-K} \frac{Y_2'(I-M)Y_2}{n}$$

or

$$\frac{Y_2'(I-M)Y_2}{n-K} = \frac{Y_2'(I-M)Y_2}{n} + \frac{K}{n-K} \frac{Y_2'(I-M)Y_2}{n}$$

so that

$$\frac{K}{n-K} \frac{Y_2'(I-M)Y_2}{n} = \frac{Y_2'(I-M)Y_2}{n-K} - \frac{Y_2'(I-M)Y_2}{n} .$$

Therefore, from this result expression 4.55 can be written

$$\hat{H} = \begin{bmatrix} \frac{Y_2' Y_2}{n} & \frac{Y_2' M X_1}{n} \\ \frac{X_1' M Y_2}{n} & \frac{X_1' X_1}{n} \end{bmatrix} - \begin{bmatrix} \frac{Y_2' (I-M) Y_2}{n-K} & 0 \\ 0 & 0 \end{bmatrix}$$

Now

$$E\left(\frac{Y_2' Y_2}{n} - \frac{Y_2' (I-M) Y_2}{n-K}\right) = E\left(\frac{\pi_2' X' X \pi_2}{n} + \frac{V_2' V_2}{n} - \frac{Y_2' (I-M) Y_2}{n-K}\right) = \pi_2' \Sigma_{XX} \pi_2$$

because  $E\left(\frac{V_2' V_2}{n}\right) = E\left(\frac{Y_2' (I-M) Y_2}{n-K}\right) = \phi_{22}$ , a submatrix of  $\phi$ . Now

$$E\left(\frac{Y_2' M X_1}{n}\right) = E\left(\frac{Y_2' X_1}{n}\right) = E\left(\frac{\pi_2' X' X_1}{n}\right) = \pi_2' \Sigma_{XX_1}$$

It follows that

$$E(\hat{H}) = \begin{bmatrix} \pi_2' \Sigma_{XX} \pi_2 & \pi_2' \Sigma_{XX_1} \\ \Sigma_{XX_1}' \pi_2 & \Sigma_{X_1 X_1} \end{bmatrix},$$

which was to be shown.

Let us consider now the following expression

$$\hat{N} = \begin{bmatrix} \frac{Y_2' M Y_1}{n} \\ \frac{X_1' Y_1}{n} \end{bmatrix} - K \begin{bmatrix} \frac{Y_2' (I-M) Y_2}{n(n-K)} \\ 0 \end{bmatrix}. \quad (4.57)$$

In the same way as we proved Theorem 7, we can show Theorem 8 given below.

Theorem 8 Let  $(X'X/n) = \Sigma_{XX}$ ,  $(X'X_1/n) = \Sigma_{XX_1}$ ,  $(X_1'X_1/n)$

$= \Sigma_{X_1X_1}$ . Then the expected value of  $\hat{N}$  given by 4.57 is

$$- \begin{bmatrix} \pi_2' \Sigma_{XX} \pi_2 & \pi_2' \Sigma_{XX_1} \\ \Sigma_{XX_1} \pi_2 & \Sigma_{X_1X_1} \end{bmatrix} \begin{bmatrix} \beta_2' \\ \gamma_1' \end{bmatrix} = N. \quad (4.58)$$

Proof: Following similar steps as in the Proof of Theorem 7,

it is easily shown that 4.57 can also be written

$$\begin{aligned} \hat{N} &= \begin{bmatrix} \frac{Y_2' Y_1}{n} \\ \frac{X_1' Y_1}{n} \end{bmatrix} - \begin{bmatrix} \frac{Y_2' (I-M) Y_1}{n-K} \\ 0 \end{bmatrix} = - \left\{ \begin{bmatrix} \frac{Y_2' Y_2}{n} & \frac{Y_2' X_1}{n} \\ \frac{X_1' Y_2}{n} & \frac{X_1' X_1}{n} \end{bmatrix} \begin{bmatrix} \beta_2' \\ \gamma_1' \end{bmatrix} \right. \\ &\quad \left. + \begin{bmatrix} \frac{Y_2' (I-M) Y_2}{n-K} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_2' \\ \gamma_1' \end{bmatrix} + \begin{bmatrix} \frac{Y_2' u_1}{n} - \frac{Y_2' (I-M) u_1}{n-K} \\ 0 \end{bmatrix} \right\} \end{aligned}$$

where we have used expression 4.42. Taking expectations we obtain the desired conclusion.

If we transform Model 4.42 by an orthogonal transformation satisfying 4.51, the relations 4.55 and 4.57 take the form

$$\hat{H} = \begin{bmatrix} \frac{Z'_{21} W_1 (W'_1 W_1)^{-1} W'_1 Z_{21}}{n} & \frac{Z'_{21} W_1}{n} \\ \frac{W'_1 Z_{21}}{n} & \frac{W'_1 W_1}{n} \end{bmatrix} - K \begin{bmatrix} \frac{Z'_1 [I - W_1 (W'_1 W_1)^{-1} W'_1] Z_{21}}{n(n-K)} & 0 \\ 0 & 0 \end{bmatrix} \quad (4.59)$$

and

$$\hat{N} = \begin{bmatrix} \frac{Z'_{21} W_1 (W'_1 W_1)^{-1} W'_1 z_{11}}{n} \\ \frac{W'_1 z_{11}}{n} \end{bmatrix} - K \begin{bmatrix} \frac{Z'_{21} [I - W_1 (W'_1 W_1)^{-1} W'_1] z_{11}}{n(n-K)} \\ 0 \end{bmatrix}, \quad (4.60)$$

where  $Z_{21}$ ,  $z_{11}$  and  $W_1 = Q_1 X$  were defined in the discussion to arrive at expressions 4.52.

Using the errors in variables form 4.52 of the simultaneous equations Model 4.42, to obtain an estimator for  $(\beta_2, \gamma_1)$  unbiased to  $O(1/n)$ , let

$$S^* = \frac{Z'_{21} [I - W_1 (W'_1 W_1)^{-1} W'_1] Z_{21}}{n(n-K)} = \frac{Y'_2 (I-M) Y_2}{n(n-K)},$$

$$R^* = \frac{Z'_{21} [I - W_1 (W'_1 W_1)^{-1} W'_1] z_{11}}{n(n-K)} = \frac{Y'_2 (I-M) y_1}{n(n-K)},$$

$$c = n\left(\frac{1}{n} + \frac{1}{n-K}\right) \approx 2 \quad \text{in large samples,}$$

$$k = G_1 + K_1 - 1.$$

Take  $\hat{H}$  and  $\hat{N}$  as the right hand sides of relations 4.59 and 4.60 and calculate

$$G^* = \{[G_1 + K_1 + \text{ctr}(\hat{H}^{11} S^*)] S^* + c(S^* \hat{H}^{11} S^*)\}, \quad (4.61)$$

and

$$D^* = \{[G_1 + K_1 + \text{ctr}(\hat{H}^{11} S^*)]R^* + c(S^* \hat{H}^{11} S^*)\} ; \quad (4.62)$$

where in 4.61 and 4.62  $\hat{H}^{11}$  is the submatrix of  $\hat{H}^{-1}$  corresponding to

$$\hat{H}_{11} = \frac{Y_2' M Y_2}{n} - K \frac{Y_2' (I-M) Y_2}{n(n-K)} ,$$

a submatrix of  $\hat{H}$ . In terms of the variables in Model 4.42, to  $O_p(1/n)$  the unbiased estimator is

$$- \begin{bmatrix} \hat{\beta}_2' \\ \tilde{y}_1' \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \{ Y_2' M Y_2 - \frac{K}{n-K} Y_2' (I-M) Y_2 \} + G^* , & \frac{1}{n} (Y_2' X_1) \\ \frac{1}{n} (X_1' Y_2) & \frac{1}{n} (X_1' X_1) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{1}{n} \{ Y_2' M Y_1 - \frac{K}{n-K} Y_2' (I-M) Y_1 \} + D^* \\ \frac{1}{n} (X_1' Y_1) \end{bmatrix} \quad (4.63)$$

where  $G^*$  and  $D^*$  are given by 4.61 and 4.62.

#### 4. Equivalence with Nagar's estimator

Nagar (1959) has obtained the unbiased member, to  $O_p(1/n)$ , of the  $K$ th class of estimators. In terms of our notation his estimator is given by

$$- \begin{bmatrix} \hat{\beta}_2' \\ \hat{y}_1' \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \{ Y_2' M Y_2 - \frac{K-(G_1+K_1)}{n} Y_2' (I-M) Y_2 \} , & \frac{1}{n} (Y_2' X_1) \\ \frac{1}{n} X_1' Y_2 & \frac{1}{n} (X_1' X_1) \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \frac{1}{n}\{Y_2'MY_1 - \frac{K-(G_1+K_1)}{n} Y_2'(I-M)y_1\} \\ \frac{1}{n} (X_1'y_1) \end{bmatrix} \quad (4.64)$$

Let us prove that 4.63 and 4.64 are equivalent except for terms of  $O_p(1/n^2)$ . By direct examination we find that 4.63 and 4.64 are the same except that

$$\frac{1}{n}\{Y_2'MY_2 - \frac{K}{n-K} Y_2'(I-M)Y_2 + nG^*\} \neq \frac{1}{n}\{Y_2'MY_2 - \frac{K-(G_1+K_1)}{n} Y_2'(I-M)Y_2\} \quad (4.65)$$

and

$$\frac{1}{n}\{Y_2'MY_1 - \frac{K}{n-K} Y_2'(I-M)y_1 + nD^*\} \neq \frac{1}{n}\{Y_2'MY_1 - \frac{K-(G_1+K_1)}{n} Y_2'(I-M)y_1\} \quad (4.66)$$

The equivalence up to terms of  $O_p(1/n)$  of the estimators 4.63 and 4.64 will be shown if we are able to prove that 4.65 and 4.66 differ only by terms of  $O_p(1/n^2)$ . By adding and subtracting

$$\frac{1}{n} \left\{ \frac{K}{n} \right\} Y_2'(I-M)Y_2$$

to the left hand side of 4.65 we obtain

$$\begin{aligned} \frac{1}{n}\{Y_2'MY_2 - \frac{K}{n-K} Y_2'(I-M)Y_2 + nG^*\} &= \frac{1}{n}\left\{Y_2'MY_2 - \frac{K-(G_1+K_1)}{n} Y_2'(I-M)Y_2\right\} \\ &+ [-K^2 + n \text{ctr}(\hat{H}^{11} S^*)] S^* + n S^* \hat{H}^{11} S^* \}. \end{aligned} \quad (4.67)$$

Assume the conditions of Theorem 7. Then the elements of  $\hat{H}$  are of  $O_p(1)$ , and of course the elements of  $\hat{H}^{-1}$  are also  $O_p(1)$ . Now  $\hat{H}^{11}$  is a submatrix of  $\hat{H}^{-1}$ , therefore its elements are  $O_p(1)$ . Also  $S^*$  is

$O_p(1/n)$  and  $c = O(1)$ . Therefore

$$- \frac{1}{n} K^2 S^* = O_p(1/n^2)$$

$$c[t_r(\hat{H}^{11} S^*)] S^* = O_p(1/n^2),$$

and

$$c S^* \hat{H}^{11} S^* = O_p(1/n^2).$$

Then,

$$\begin{aligned} \frac{1}{n} \{ Y_2' M Y_2 - \frac{K}{n-K} Y_2' (I-M) Y_2 + n G^* \} &= \frac{1}{n} \{ Y_2' M Y_2 - \frac{K-(G_1+K_1)}{n} Y_2' (I-M) Y_2 \} \\ &+ O_p(1/n^2). \end{aligned} \quad (4.68)$$

In exactly the same way it is possible to show

$$\begin{aligned} \frac{1}{n} \{ Y_2' M Y_1 - \frac{K}{n-K} Y_2' (I-M) Y_1 + n D^* \} &= \frac{1}{n} \{ Y_2' M Y_1 - \frac{K-(G_1+K_1)}{n} Y_2' (I-M) Y_1 \} \\ &+ O_p(1/n^2). \end{aligned} \quad (4.69)$$

Let  $\tilde{H}^{-1} \tilde{N}$  denote Nagar's estimator, where  $\tilde{H}$  is the first term on the right hand side of 4.68 and  $\tilde{N}$  the first term on the right hand side of 4.69.

Our estimator in 4.63 can be written

$$\begin{aligned} - \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_1' \end{bmatrix} &= [\tilde{H} + O_p(1/n^2)]^{-1} [\tilde{N} + O_p(1/n^2)] \\ &= [\tilde{H} (I - \tilde{H}^{-1} O_p(1/n^2))]^{-1} [\tilde{N} + O_p(1/n^2)] \\ &= [\tilde{H}^{-1} - \tilde{H}^{-1} O_p(1/n^2) \tilde{H}^{-1} + \dots] [\tilde{N} + O_p(1/n^2)] \end{aligned}$$

$$\begin{aligned}
&= \tilde{H}^{-1} \tilde{N} + \tilde{H}^{-1} O_p(1/n^2) - \tilde{H}^{-1} O_p(1/n^2) \tilde{H}^{-1} \tilde{N} + \dots \\
&= \tilde{H}^{-1} \tilde{N} + O_p(1/n^2)
\end{aligned} \tag{4.70}$$

Expression 4.70 establishes the equivalence of Nagar's estimator and the estimator given by 4.63, except for terms of  $O_p(1/n^2)$ .

#### D. Nearly Unbiased Estimators of Certain Functions of the Parameters of a Regression Model

Let us consider the multiple regression model

$$y = X\beta + e \tag{4.71}$$

where  $y$  is the  $n \times 1$  vector of observations,  $X$  a  $n \times p$  ( $n \geq p$ ) matrix of fixed constants,  $e$  is the  $n \times 1$  vector of errors and  $\beta$  is the  $p \times 1$  vector of unknown parameters. Let  $e$  be distributed according to a  $n$ -variate normal with mean zero and covariance matrix  $I\sigma^2$ , where  $I$  is the  $n \times n$  identity matrix. Then we know that the minimum variance unbiased estimator of  $\beta$  is given by

$$\hat{\beta} = (X'X)^{-1}X'y \tag{4.72}$$

From the above assumptions, it follows that  $\hat{\beta}$  has a  $p$ -variate normal distribution with mean  $\beta$  and covariance matrix  $(X'X)^{-1}\sigma^2$ . Let  $\frac{1}{n}(X'X) = O(1)$ .

Let us write

$$\beta' = (\beta_1, \beta_2, \dots, \beta_p) \tag{4.73}$$

and

$$\hat{\beta}' = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p) \tag{4.74}$$



We shall consider the estimation of the function

$$C = H^{-1}N, \quad (4.74)$$

where

$$H = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kk} \end{bmatrix} \quad (4.75)$$

and

$$N' = (\beta_{10} \ \beta_{20} \ \cdots \ \beta_{k0}) \ . \quad (4.76)$$

The vectors  $(\beta_{1j}, \beta_{2j}, \dots, \beta_{kj})'$ ,  $j = 0, 1, \dots, k$ ,  $k < p$ , entering in 4.75 and 4.76 are arbitrary permutations of size  $k$  of the elements of  $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ . Consider

$$\hat{C} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \cdots & \hat{\beta}_{1k} \\ \hat{\beta}_{21} & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \hat{\beta}_{k1} & \hat{\beta}_{k2} & \cdots & \hat{\beta}_{kk} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_{10} \\ \hat{\beta}_{20} \\ \cdot \\ \hat{\beta}_{k0} \end{bmatrix} \quad (4.77)$$

where the vectors  $(\hat{\beta}_{1j}, \hat{\beta}_{2j}, \dots, \hat{\beta}_{kj})'$ ,  $j = 0, 1, \dots, k$ ,  $k \leq p$ , are the corresponding permutations of size  $k$  of the elements of  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)'$ .

Another way of writing 4.77 is

$$\hat{C} = (H+a)^{-1}(N+b) = \hat{H}^{-1}\hat{N} \ , \quad (4.78)$$

where  $H$  and  $N$  are given by 4.75 and 4.76, respectively, and

$$a = \begin{bmatrix} \hat{\beta}_{11} - \beta_{11} & \hat{\beta}_{12} - \beta_{12} & \cdots & \hat{\beta}_{1k} - \beta_{1k} \\ \hat{\beta}_{21} - \beta_{21} & \hat{\beta}_{22} - \beta_{22} & \cdots & \hat{\beta}_{2k} - \beta_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \hat{\beta}_{k1} - \beta_{k1} & \hat{\beta}_{k2} - \beta_{k2} & \cdots & \hat{\beta}_{kk} - \beta_{kk} \end{bmatrix} = O_p(1/\sqrt{n}), \quad (4.79)$$

and

$$b' = (\hat{\beta}_{10} - \beta_{10} \quad \hat{\beta}_{20} - \beta_{20} \quad \cdots \quad \hat{\beta}_{k0} - \beta_{k0}) = O_p(1/\sqrt{n}). \quad (4.80)$$

Proceeding as in Section A of this chapter, we have to  $O_p(1/n)$

$$\bar{E}(\hat{C}) - C = H^{-1}\{[E(aH^{-1}a)]C - E(aH^{-1}b)\}. \quad (4.81)$$

Let  $G = E(aH^{-1}a)$ , and let  $G_{ij}$  be the element on the  $i$ th row,  $j$ th column of  $G$ . Let  $\phi_{qr}^{(ij)} = E\{(\hat{\beta}_{iq} - \beta_{iq})(\hat{\beta}_{rj} - \beta_{rj})\}$ ,  $1 \leq i, j \leq k$ ,  $i \leq q, r \leq k$ . Define the  $k \times k$  matrix  $\phi^{(ij)}$  as follows

$$\phi^{(ij)} = \begin{bmatrix} \phi_{11}^{(ij)} & \phi_{12}^{(ij)} & \cdots & \phi_{1k}^{(ij)} \\ \phi_{21}^{(ij)} & \phi_{22}^{(ij)} & \cdots & \phi_{2k}^{(ij)} \\ \cdot & \cdot & \cdot & \cdot \\ \phi_{k1}^{(ij)} & \phi_{k2}^{(ij)} & \cdots & \phi_{kk}^{(ij)} \end{bmatrix}, \quad 1 \leq i, j \leq k. \quad (4.82)$$

There will be in general  $k^2$  such matrices. To construct these matrices, we take the necessary elements from the matrix  $(X'X)^{-1}\sigma^2$ . Let  $H^{ij}$  be the element on the  $i$ th row,  $j$ th column of  $H^{-1}$ . Then, it is easily shown that

$$E(G_{ij}) = \sum_{q=1}^k \sum_{r=1}^k H^{qr} \phi_{qr}^{(ij)} = \text{tr} (H^{-1} \phi^{(ij)}) , \quad (4.83)$$

$$1 \leq i, j \leq k .$$

Similarly, let  $\phi_{qr}^{(i)} = E\{(\hat{\beta}_{iq} - \beta_{iq})(\hat{\beta}_{r0} - \beta_{r0})\}$ ,  $1 \leq i, q, r \leq k$ . For every  $i$  we can construct the matrix

$$\phi^{(i)} = \begin{bmatrix} \phi_{11}^{(i)} & \phi_{12}^{(i)} & \dots & \phi_{1k}^{(i)} \\ \phi_{21}^{(i)} & \phi_{22}^{(i)} & \dots & \phi_{2k}^{(i)} \\ \dots & \dots & \dots & \dots \\ \phi_{k1}^{(i)} & \phi_{k2}^{(i)} & \dots & \phi_{kk}^{(i)} \end{bmatrix} , \quad 1 \leq i \leq k . \quad (4.84)$$

If we write  $D = E(aH^{-1}b)$ , and let  $D_i$  be the  $i$ th element of the vector  $D$ , then, it can be shown that

$$E(D_i) = \sum_{q=1}^k \sum_{r=1}^k H^{qr} \phi_{qr}^{(i)} = \text{tr} (H^{-1} \phi^{(i)}) , \quad (4.85)$$

$$1 \leq i \leq k .$$

By substitution of 4.83 and 4.85 in 4.81 we can obtain an expression for the bias to  $O_p(1/n)$ .

Now, if we set

$$\hat{G}_{ij} = \text{tr} (\hat{H}^{-1} \hat{\phi}^{(ij)}) , \quad (4.86)$$

and

$$\hat{D}_i = \text{tr} (\hat{H}^{-1} \hat{\phi}^{(i)}) , \quad (4.87)$$

where  $\hat{H}$ ,  $\hat{\phi}^{(ij)}$  and  $\hat{\phi}^{(i)}$  are the simple least squares estimates of  $H$ ,  $\phi^{(ij)}$  and  $\phi^{(i)}$ . Then

$$\tilde{C} = (\hat{H} + \hat{G})^{-1} (\hat{N} + \hat{D}) \quad (4.88)$$

is an estimator of  $C$ , unbiased to  $O_p(1/n)$ . The matrices  $\hat{H}$ ,  $\hat{\phi}^{(ij)}$  and  $\hat{\phi}^{(i)}$  are constructed by taking the necessary elements from  $\hat{\beta}' = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  and from  $(X'X)^{-1}s^2$ , where  $s^2$  is the least squares estimate of  $\sigma^2$ .

$\hat{C}$  given by 4.78 does not have in general finite first moments. For example, when  $\hat{C}$  is a scalar,  $\hat{C}$  is the ratio of two normal variables. In this case,  $E|\hat{C}|$  does not exist. To prove the latter assertion, it is sufficient to prove that  $E|\hat{H}^{-1}|$  does not exist. We shall prove the following result:

Theorem 9 Let  $Z$  be a random variable having the univariate normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Then

$$E\left(\frac{1}{Z}\right) = \int_{-\infty}^{\infty} \left|\frac{1}{z}\right| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz \quad (4.89)$$

diverges.

Proof: Let  $R$  be some finite positive real number.

To show that

$$\lim_{R \rightarrow \infty} \int_{-R}^R \left|\frac{1}{z}\right| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

diverges it will be sufficient to show that

$$\int_0^R \left|\frac{1}{z}\right| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz \quad (4.90)$$

diverges for every positive  $R$ . Consider the transformation  $Z = 1/X$ , then  $dz = (-1) x dx$ . Substituting in 4.90 we obtain

$$\int_{\infty}^{1/R} (-1)^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{1}{x} - \mu)^2}{2\sigma^2}} x^{-2} dx = \int_{1/R}^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\frac{1}{x} - \mu)^2}{2\sigma^2}} dx \quad (4.91)$$

Now:

$$(i) \quad f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\frac{1}{x} - \mu)^2}{2\sigma^2}} \quad \frac{1}{R} \leq x < \infty \quad (4.92)$$

is continuous in  $\frac{1}{R} \leq x < \infty$ .

$$(ii) \quad \lim_{x \rightarrow \infty} xf(x) = \lim_{x \rightarrow \infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{1}{x} - \mu)^2}{2\sigma^2}} \quad (4.93)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}}$$

$$> 0,$$

is a finite real number. Let us now state Widder's (1961) Theorem 5, page 330:

(1)  $f(x)$  continuous  $a \leq x < \infty$ , and

(2)  $\lim_{x \rightarrow \infty} xf(x) = A \neq 0$  (or  $= \pm \infty$ )

$$\Rightarrow \int_a^{\infty} f(x) dx \text{ diverges.}$$

4.92 and 4.93 are equivalent to the hypothesis of Widder's theorem.

Therefore 4.91 and consequently 4.89 diverge.

Applying the theory developed in Section B of this chapter, it is easily seen that an estimator for  $C$  unbiased to  $O_p(1/n)$ , with finite first moments can be constructed. Let

$$\tilde{H} = (\hat{H}'\hat{H} + P)^{-1}\hat{H}' + (\hat{H}'\hat{H} + P)^{-1}P(\hat{H}'\hat{H} + P)^{-1}\hat{H}'$$

where

$$P = \begin{bmatrix} \hat{\phi}_{11}^{(11)} & \hat{\phi}_{12}^{(12)} & \dots & \hat{\phi}_{1k}^{(1k)} \\ \hat{\phi}_{21}^{(21)} & \hat{\phi}_{22}^{(22)} & \dots & \hat{\phi}_{2k}^{(2k)} \\ \dots & \dots & \dots & \dots \\ \hat{\phi}_{k1}^{(k1)} & \hat{\phi}_{k2}^{(k2)} & \dots & \hat{\phi}_{kk}^{(kk)} \end{bmatrix} ;$$

$\hat{\phi}_{ij}^{(ij)} = \hat{E}(\hat{\beta}_{ii} - \beta_{ii})(\hat{\beta}_{jj} - \beta_{jj})$ ,  $1 \leq i, j \leq k$ . Then using expressions 4.38 and 4.39,  $\hat{H} = \tilde{H} + O_p(1/n^2)$ . Also, if  $(1/n)(X'X) = O(1)$ , the elements of  $P$  will be  $O_p(1/n)$ . Therefore, the estimator

$$\tilde{C} = (\tilde{H} + \hat{G})^{-1}(\hat{N} + \hat{D})$$

will have finite  $r$ th moments, provided  $n \geq 2r+k+2$ , and will be unbiased for  $C$ , to  $O_p(1/n)$ .

## V. ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR

Due to the absence of knowledge about the properties of the maximum likelihood estimator for the errors in variables model, some research was needed on this area. We present in this chapter the author's results on the errors in variables model, one variable case. Further developments for the several variables case obtained by Fuller (1970b) are given. Some few remarks are presented with respect to the efficiency of the maximum likelihood estimators. The subject is introduced describing the maximum likelihood estimation procedure for the errors in variables model.

### A. Errors in Variables, Classical Case

Let us rewrite the errors in variables model in a slightly different form

$$\sum_{i=0}^k \alpha_i x_{ij} = 0, \quad j = 1, 2, \dots, n,$$

where  $x_{0j} = 1$ ,  $j = 1, 2, \dots, n$ . In matrix notation we shall have

$$\alpha_0 + \alpha' x_j = 0, \quad j = 1, 2, \dots, n, \quad (5.1)$$

where  $\alpha$  and  $x_j$  are  $k \times 1$  vectors. We observe  $X_{ij}$  such that

$$X_{ij} = x_{ij} + u_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n,$$

or in matrix notation

$$X_j = x_j + u_j, \quad j = 1, \dots, n.$$

Assume

- (i)  $u_j$ ,  $j = 1, 2, \dots, n$ , is independently distributed according to  $N_k(0, V\sigma^2)$ .
- (ii)  $V$ ,  $k \times k$  is known.
- (iii) The  $x_{ij}$ 's are fixed unknown constants.

The  $\alpha$ 's,  $x$ 's and  $\sigma^2$  are parameters to be estimated. The likelihood function is

$$L = (\text{Constant}) \left| V\sigma^2 \right|^{-\frac{n}{2}} e^{-\frac{1}{2} \left\{ \sum_{j=1}^n u_j' (V\sigma^2)^{-1} u_j \right\}} \quad (5.2)$$

Our problem is to find the values of the  $\alpha$ 's and the  $x$ 's which maximize 5.2 subject to the restrictions 5.1. Shifting to deviations, the problem takes the following form (see Johnston (1963) page 171). Let

$$w_j = \begin{bmatrix} (x_{1j} - \bar{x}_1) - (x_{1j} - \bar{x}_1) \\ \vdots \\ (x_{kj} - \bar{x}_k) - (x_{kj} - \bar{x}_k) \end{bmatrix}, \quad j = 1, 2, \dots, n.$$

Maximizing  $L$  is equivalent to minimizing

$$\sum_{j=1}^n w_j' V^{-1} w_j, \quad ,$$

subject to the restrictions

$$\sum_{i=1}^k \alpha_i (x_{ij} - \bar{x}_i) = \alpha' (x_j - \bar{x}) = 0, \quad j = 1, \dots, n. \quad (5.3)$$

We form the function



$$\psi = \sum_{j=1}^n w_j' V^{-1} w_j + \sum_{j=1}^n \mu_j \alpha' (x_j - \bar{x}) , \quad (5.4)$$

where the  $\mu$ 's are Lagrange's multipliers. From 5.4 we obtain

$$\frac{\partial \psi}{\partial (x_j - \bar{x})} = -2V^{-1} w_j + \mu_j \alpha = 0 , \quad j = 1, \dots, n . \quad (5.5)$$

$$w_j = \frac{1}{2} \mu_j V \alpha , \quad j = 1, \dots, n . \quad (5.6)$$

Multiplying 5.6 on the left by  $\alpha'$ , we obtain

$$\alpha' w_j = \frac{1}{2} \mu_j \alpha' V \alpha , \quad j = 1, \dots, n . \quad (5.7)$$

Multiplying 5.5 on the left by  $w_j'$ , we obtain

$$-2w_j' V^{-1} w_j + \mu_j w_j' \alpha = 0 , \quad j = 1, \dots, n . \quad (5.8)$$

From 5.7

$$\frac{1}{2} \mu_j = \frac{\alpha' w_j}{\alpha' V \alpha} , \quad j = 1, \dots, n ,$$

which placed in 5.8 gives

$$-w_j' V^{-1} w_j + \frac{\alpha' w_j w_j' \alpha}{\alpha' V \alpha} = 0 , \quad j = 1, \dots, n . \quad (5.9)$$

Therefore summing over  $j$  in 5.9

$$\begin{aligned} \sum_{j=1}^n w_j' V^{-1} w_j &= \frac{\alpha' \left( \sum_{j=1}^n w_j w_j' \right) \alpha}{\alpha' V \alpha} \\ &= \frac{\alpha' \left\{ \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x}) \right\} \alpha}{\alpha' V \alpha} . \end{aligned} \quad (5.10)$$

To obtain expression 5.10 we used the restrictions 5.3; that is,  $\alpha'(x_j - \bar{x}) = 0$  for all  $j$ . Clearly, minimizing  $\sum_{j=1}^n w_j' V^{-1} w_j$  is the same as minimizing  $\alpha' M \alpha / \alpha' V \alpha$ , where

$$M = \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' .$$

Or identically, the problem reduces to minimizing  $\alpha' M \alpha$  subject to the restriction  $\alpha' V \alpha = C$ . We form the function

$$\phi = \alpha' M \alpha - \lambda(\alpha' V \alpha - C) , \quad (5.11)$$

where  $\lambda$  is Lagrange's multiplier. From 5.11 we get

$$\frac{\partial \phi}{\partial \alpha} = 2M\alpha - 2\lambda V\alpha = 0 ,$$

$$\alpha' V \alpha - C = 0 .$$

We normalize by setting  $C = 1$ . Then, the first of the above equations gives

$$(M - \lambda V)\alpha = 0 , \quad (5.12)$$

but the above system has a nontrivial solution if

$$|M - \lambda V| = 0 .$$

The minimum root of the last expression will make the likelihood function a maximum. To prove it, we shall use the theorem given in Bodewig (1956) page 65, used in Section B of Chapter IV to prove Lemma 2. Part 3 of the theorem states that "the smallest eigenvalue of the real symmetric matrix  $A$ , is the minimum of the quadratic form  $x'Ax$  for all vectors  $x$  such that  $x'x = 1$ ." In our case,  $V$  positive definite symmetric, implies

the existence of a nonsingular matrix  $T$ ,  $k \times k$ , such that  $V = T'T$ . By substitution in 5.12 we obtain

$$(M - \lambda T'T)\alpha = 0 \quad .$$

From the above expression, we obtain

$$(T'^{-1}MT^{-1} - \lambda I)T\alpha = 0 \quad .$$

Let  $\beta = T\alpha$ . By the theorem in Bodewig

$$\begin{aligned} \lambda_1 &= \min_{\{\beta: \beta'\beta=1\}} \frac{\beta'T'^{-1}MT^{-1}\beta}{\beta'\beta} \\ &= \text{minimum characterist root of } T'^{-1}MT^{-1} \\ &= \min_{\{\alpha: \alpha'V\alpha=1\}} \frac{\alpha'M\alpha}{\alpha'V\alpha} \\ &= \text{minimum root of system 5.12} \quad . \end{aligned}$$

Then

$$(M - \lambda_1 V)\hat{\alpha} = 0 \quad , \quad (5.13)$$

will give the maximum likelihood estimator of the vector  $\alpha$  under conditions (i) - (iii), stated above.

When  $V$  is not known but there is available a function  $S$ , independent of the errors  $u_j$ , such that

$$E(S) = E(u_j u_j') \quad , \quad j = 1, \dots, n \quad ,$$

we could substitute  $S$  for  $V$  in the above procedure. There is a loss in efficiency relative to the case when  $V$  is known. We shall call both

types of estimators "the maximum likelihood estimators" in the following pages.

B. The Asymptotic Bias and Variance of the Maximum Likelihood Estimators. Errors in Variables, One Variable Case

There has been little work with respect to the asymptotic properties of the maximum likelihood estimators. Malinvaud (1966) has given a solution for the asymptotic variance to  $O(1/n)$  when  $V$  is known. However, his solution is implicitly stated in a system of simultaneous equations. To the author's knowledge no work has been reported with respect to the asymptotic bias of the maximum likelihood estimators. Using the Taylor's expansions discussed in Chapter III, we shall establish in this section the asymptotic bias and variance of the maximum likelihood estimator. We shall consider the one variable case,  $V$  unknown.

In the one variable case we have

$$\begin{aligned} Y_j &= y_j + e_j, \quad j = 1, \dots, n, \\ y_j &= x_j \beta, \\ X_j &= x_j + u_j, \end{aligned} \tag{5.14}$$

where  $(e_j, u_j)'$ ,  $j = 1, \dots, n$ , is independent bivariate normal with zero means and

$$E(e_j, u_j)'(e_j, u_j) = \begin{bmatrix} \sigma_e^2 & \sigma_{eu} \\ \sigma_{ue} & \sigma_u^2 \end{bmatrix}, \quad j = 1, \dots, n; \quad \sigma_{ue} = \sigma_{eu}.$$

The y's and the x's are fixed unknown constants. Assume we have a matrix

$$S = \begin{bmatrix} s_e^2 & s_{eu} \\ s_{ue} & s_u^2 \end{bmatrix},$$

independent of the errors, such that  $S$  is Wishart with  $d$  degrees of freedom. Moreover,  $S$  satisfies

$$E(S) = \begin{bmatrix} \sigma_e^2 & \sigma_{eu} \\ \sigma_{ue} & \sigma_u^2 \end{bmatrix}.$$

Let

$$M = \begin{bmatrix} m_{YY} & m_{YX} \\ m_{XY} & m_{XX} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n \frac{1}{n} (Y_j - \bar{Y})^2 & \sum_{j=1}^n \frac{1}{n} (Y_j - \bar{Y})(X_j - \bar{X}) \\ \sum_{j=1}^n \frac{1}{n} (X_j - \bar{X})(Y_j - \bar{Y}) & \sum_{j=1}^n \frac{1}{n} (X_j - \bar{X})^2 \end{bmatrix}.$$

We assume that

$$E(M-S) = \begin{bmatrix} \sum_j \frac{1}{n} (y_j - \bar{y})^2 & \sum_j \frac{1}{n} (y_j - \bar{y})(x_j - \bar{x}) \\ \sum_j \frac{1}{n} (x_j - \bar{x})(y_j - \bar{y}) & \sum_j \frac{1}{n} (x_j - \bar{x})^2 \end{bmatrix} \rightarrow \begin{bmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_x^2 \end{bmatrix},$$

fixed, as  $n \rightarrow \infty$ , say. The vector  $\alpha' = (1, -\beta)$ .

If  $\lambda_1$  is the minimum root of

$$\left| \begin{bmatrix} m_{YY} & m_{YX} \\ m_{XY} & m_{XX} \end{bmatrix} - \lambda \begin{bmatrix} s_e^2 & s_{eu} \\ s_{ue} & s_u^2 \end{bmatrix} \right| = 0,$$

Then, the maximum likelihood estimator of  $\beta$  is given by

$$\left[ \begin{pmatrix} m_{YY} & m_{YX} \\ m_{XY} & m_{XX} \end{pmatrix} - \lambda_1 \begin{pmatrix} s_e^2 & s_{eu} \\ s_{ue} & s_u^2 \end{pmatrix} \right] \begin{bmatrix} 1 \\ -\hat{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

It is easily seen that  $\lambda_1$  is the minimum root of

$$\begin{aligned} (m_{XX}m_{YY} - m_{XY}^2) - \lambda(m_{XX}s_e^2 + m_{YY}s_u^2 - 2m_{XY}s_{ue}) \\ + \lambda^2(s_e^2s_u^2 - s_{ue}^2) = 0 \end{aligned}$$

$$\lambda_1 = \frac{1}{2(s_e^2s_u^2 - s_{ue}^2)} \{ (m_{XX}s_e^2 + m_{YY}s_u^2 - 2m_{XY}s_{ue})$$

$$- [(m_{XX}s_e^2 + m_{YY}s_u^2 - 2m_{XY}s_{ue})^2 - 4(s_e^2s_u^2 - s_{ue}^2)(m_{XX}m_{YY} - m_{XY}^2)]^{\frac{1}{2}} \} .$$

Therefore,

$$\begin{aligned} \hat{\beta} &= \frac{m_{XY} - \lambda_1 s_{ue}}{m_{XX} - \lambda_1 s_u^2} = \frac{G}{H} \\ &= \frac{B + s_{ue} A^{\frac{1}{2}}}{C + s_u^2 A^{\frac{1}{2}}} ; \end{aligned} \quad (5.15)$$

where

$$B = 2m_{XY}(s_u^2s_e^2 - s_{ue}^2) - s_{ue}(s_u^2m_{YY} - 2s_{ue}m_{XY} + s_e^2m_{XX}) ;$$

$$C = 2m_{XX}(s_u^2s_e^2 - s_{ue}^2) - s_u^2(s_u^2m_{YY} - 2s_{ue}m_{XY} + s_e^2m_{XX}) ;$$

$$A = (s_u^2m_{YY} - 2s_{ue}m_{XY} + s_e^2m_{XX})^2 - 4(s_u^2s_e^2 - s_{ue}^2)(m_{XX}m_{YY} - m_{XY}^2) .$$

Define

$$a = s_u^2 s_e^2 - s_{ue}^2 ,$$

$$b = s_u^2 m_{YY} - 2s_{ue} m_{XY} + s_e^2 m_{XX} ,$$

$$c = m_{XX} m_{YY} - m_{XY}^2 .$$

Now,

$$a \xrightarrow{P} \bar{a} = \sigma_u^2 \sigma_e^2 - \sigma_{ue}^2 ,$$

$$b \xrightarrow{P} \bar{b} = \sigma_x^2 (\beta^2 \sigma_u^2 - 2\beta \sigma_{ue} + \sigma_e^2) + 2(\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2) ,$$

$$c \xrightarrow{P} \bar{c} = \sigma_x^2 (\beta^2 \sigma_u^2 - 2\beta \sigma_{ue} + \sigma_e^2) + (\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2) ,$$

$$A \xrightarrow{P} \bar{A} = \bar{b}^2 - 4\bar{a}\bar{c} = (\beta^2 \sigma_u^2 - 2\beta \sigma_{ue} + \sigma_e^2)^2 = (\bar{c} - \bar{a})^2 ,$$

$$G \xrightarrow{P} \bar{G} = 2\beta (\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2) \sigma_x^2 ,$$

$$H \xrightarrow{P} \bar{H} = 2 (\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2) \sigma_x^2 .$$

Let us write for the moment  $z_1 = s_u^2$  ,  $z_2 = s_e^2$  ,  $z_3 = s_{ue}$  ,  $z_4 = m_{XX}$  ,  $z_5 = m_{YY}$  ,  $z_6 = m_{XY}$  . If  $k_i$  ,  $i = 1, \dots, 6$  , denotes the probability limit of  $z_i$  ,  $i = 1, \dots, 6$  , it is easily shown that

$$\hat{\beta} - \beta = \bar{H}^{-1} \left\{ \sum_i \left[ \frac{\partial G}{\partial z_i} - \beta \frac{\partial H}{\partial z_i} \right] (z_i - k_i) \right\} + o_p(1/n) ,$$

where  $k = (k_1, \dots, k_6)$  . Thus, to  $O(1/n)$

$$\begin{aligned} E(\hat{\beta} - \beta)^2 &= \bar{H}^{-2} \sum_{ij} \left[ \frac{\partial G}{\partial z_i} - \beta \frac{\partial H}{\partial z_i} \right] \left[ \frac{\partial G}{\partial z_j} - \beta \frac{\partial H}{\partial z_j} \right] \\ &\quad \cdot E(z_i - k_i)(z_j - k_j) . \end{aligned} \quad (5.16)$$

5.16 gives to  $o_p(1/n)$  the asymptotic variance of  $\hat{\beta}$ .

To obtain an expression for the asymptotic variance of  $\hat{\beta}$  we shall also need the following results

$$s_u^2 \xrightarrow{P} \sigma_u^2 ,$$

$$s_e^2 \xrightarrow{P} \sigma_e^2 ,$$

$$s_{ue}^2 \xrightarrow{P} \sigma_{ue} ,$$

$$m_{XX} \xrightarrow{P} \sigma_x^2 + \sigma_u^2$$

$$m_{YY} \xrightarrow{P} \beta^2 \sigma_x^2 + \sigma_e^2$$

$$m_{XY} \xrightarrow{P} \beta \sigma_x^2 + \sigma_{ue} .$$

Also

$$\text{Var} (s_u^2) = (1/d) 2 (\sigma_u^2)^2 ,$$

$$\text{Var} (s_e^2) = (1/d) 2 (\sigma_e^2)^2 ,$$

$$\text{Var} (s_{ue}) = (1/d)(\sigma_u^2 \sigma_e^2 + \sigma_{ue}^2) ,$$

$$\text{Cov} (s_u^2, s_e^2) = (1/d) 2 \sigma_{ue}^2 ,$$

$$\text{Cov} (s_u^2, s_{ue}) = (1/d) 2 \sigma_u^2 \sigma_{ue} ,$$

$$\text{Cov} (s_e^2, s_{ue}) = (1/d) 2 \sigma_e^2 \sigma_{ue} ,$$

where  $d$  is the degrees of freedom associated with the  $S$  . Similarly

$$\text{Var} (m_{XX}) = \frac{1}{n-1} [4\sigma_x^2 \sigma_u^2 + 2(\sigma_u^2)^2] ,$$

$$\text{Var} (m_{YY}) = \frac{1}{n-1} [4\beta^2 \sigma_x^2 \sigma_e^2 + 2(\sigma_e^2)^2] ,$$

$$\text{Var} (m_{XY}) = \frac{1}{n-1} [\beta^2 \sigma_x^2 \sigma_u^2 + \sigma_x^2 \sigma_e^2 + 2\beta \sigma_x^2 \sigma_{ue} + \sigma_{ue}^2 + \sigma_u^2 \sigma_e^2] ,$$



$$\text{Cov } (m_{XX}, m_{YY}) = \frac{1}{n-1} [4\beta\sigma_x^2\sigma_{ue}^2 + 2\sigma_{ue}^2] ,$$

$$\text{Cov } (m_{XX}, m_{XY}) = \frac{1}{n-1} [2\sigma_x^2\sigma_{ue}^2 + 2\beta\sigma_x^2\sigma_u^2 + 2\sigma_u^2\sigma_{ue}^2] ,$$

$$\text{Cov } (m_{YY}, m_{XY}) = \frac{1}{n-1} [2\beta\sigma_x^2\sigma_e^2 + 2\beta^2\sigma_x^2\sigma_{ue}^2 + 2\sigma_e^2\sigma_{ue}^2] .$$

Using the above notation, after some algebra we obtain

$$\left. \frac{\partial G}{\partial s_u^2} - \beta \frac{\partial H}{\partial s_u^2} \right|_k = (1 - \bar{b}\bar{A}^{-\frac{1}{2}})\phi \cdot \beta^2\sigma_x^2 + 2\bar{a}\bar{\beta} ,$$

$$\left. \frac{\partial G}{\partial s_e^2} - \beta \frac{\partial H}{\partial s_e^2} \right|_k = (1 - \bar{b}\bar{A}^{-\frac{1}{2}})\phi \cdot \sigma_x^2 ,$$

$$\left. \frac{\partial G}{\partial s_{ue}^2} - \beta \frac{\partial H}{\partial s_{ue}^2} \right|_k = -2(1 - \bar{b}\bar{A}^{-\frac{1}{2}})\phi \cdot \beta\sigma_x^2 - 2\bar{a} ,$$

$$\left. \frac{\partial G}{\partial m_{XX}} - \beta \frac{\partial H}{\partial m_{XX}} \right|_k = -2\bar{a}\bar{\beta} + 2\bar{a}\bar{A}^{-\frac{1}{2}}\phi \cdot \beta^2\sigma_x^2 ,$$

$$\left. \frac{\partial G}{\partial m_{YY}} - \beta \frac{\partial H}{\partial m_{YY}} \right|_k = 2\bar{a}\bar{A}^{-\frac{1}{2}}\phi \cdot \sigma_x^2 ,$$

$$\left. \frac{\partial G}{\partial m_{XY}} - \beta \frac{\partial H}{\partial m_{XY}} \right|_k = 2\bar{a} - 4\bar{a}\bar{A}^{-\frac{1}{2}}\phi \cdot \beta\sigma_x^2 ,$$

where  $\phi = \beta\sigma_u^2 - \sigma_{ue}^2$ ; the rest of the symbols having been defined.

Placing the above results in 5.16 we obtain

$$\bar{E}(\hat{\beta} - \beta)^2 = \frac{1}{(\sigma_x^2)^2} \left\{ \frac{(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2) + \sigma_x^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2)}{n-1} + \frac{(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2)}{d} \right\} \quad (5.17)$$

where we have included terms of  $O_p(1/n)$ . Letting  $d \rightarrow +\infty$  in 5.17 we obtain the expression for the variance when  $V$  is known. That is, the case for which we know  $E(e_j, u_j)'(e_j, u_j)$ ,  $j = 1, \dots, n$  except for a factor of proportionality. Let  $\tilde{\beta}$  be the estimator when  $V$  is known. Thus, the

relative efficiency of  $\hat{\beta}$  to  $\tilde{\beta}$  is, therefore,

$$\text{R.E.} = \frac{1}{1 + \frac{(n-1)(\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2)}{d[(\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2) + \sigma_x^2(\beta^2 \sigma_u^2 - 2\beta \sigma_{ue} + \sigma_e^2)]}} \quad (5.18)$$

From 5.18 we conclude that  $\tilde{\beta}$  is always more efficient than  $\hat{\beta}$ .

Using the same approach, it can be shown that

$$\bar{E}(\hat{\beta} - \beta) = \frac{1}{2}(W+N) + O_p(n^{-3/2}) \quad , \quad (5.19)$$

where

$$W = \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial^2 \hat{\beta}}{\partial z_i \partial z_j} \Big|_k \text{Cov}(z_i, z_j)$$

$$N = \sum_{i=4}^6 \sum_{j=4}^6 \frac{\partial^2 \hat{\beta}}{\partial z_i \partial z_j} \Big|_k \text{Cov}(z_i, z_j) \quad .$$

Writing in more detail we have successively:

$$W = \bar{H}^{-2} \{F + \bar{H}D\} \quad ,$$

$$D = D_1 + D_2 + D_3 \quad ,$$

$$D_1 = (-\beta A^{-\frac{1}{2}} \frac{\partial A}{\partial s_u^2}) \text{Var}(s_u^2) + (-\frac{\beta}{2} A^{-\frac{1}{2}} \frac{\partial A}{\partial s_e^2}) \cdot 2 \text{Cov}(s_u^2, s_e^2)$$

$$+ \frac{1}{2} A^{-\frac{1}{2}} (\frac{\partial A}{\partial s_{ue}} - \beta \frac{\partial A}{\partial s_u^2}) \cdot 2 \text{Cov}(s_u^2, s_{ue}) + \frac{1}{2} A^{-\frac{1}{2}} \frac{\partial A}{\partial s_e^2} \cdot 2 \text{Cov}(s_e^2, s_{ue})$$

$$+ A^{-\frac{1}{2}} \frac{\partial A}{\partial s_{ue}} \cdot \text{Var}(s_{ue}) \quad ,$$

$$\begin{aligned}
D_2 = & -\beta \frac{\partial^2 C}{\partial (s_u^2)^2} \text{Var}(s_u^2) + \left( \frac{\partial^2 B}{\partial s_u^2 \partial s_e^2} - \beta \frac{\partial^2 C}{\partial s_u^2 \partial s_e^2} \right) \cdot 2 \text{Cov}(s_u^2, s_e^2) \\
& + \left( \frac{\partial^2 B}{\partial s_u^2 \partial s_e^2} - \beta \frac{\partial^2 C}{\partial s_u^2 \partial s_e^2} \right) \cdot 2 \text{Cov}(s_u^2, s_{ue}) + \frac{\partial^2 B}{\partial s_e^2 \partial s_{ue}} \cdot 2 \text{Cov}(s_e^2, s_{ue}) \\
& - \beta \frac{\partial^2 C}{\partial s_{ue}^2} \text{Var}(s_{ue}) ,
\end{aligned}$$

$$\begin{aligned}
D_3 = & (\beta s_u^2 s_{ue}) \left\{ \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial (s_u^2)^2} + \frac{1}{4} A^{-3/2} \left( \frac{\partial A}{\partial s_u} \right)^2 \right] \text{Var}(s_u^2) \right. \\
& + \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial s_u^2 \partial s_e^2} + \frac{1}{4} A^{-3/2} \frac{\partial A}{\partial s_u} \frac{\partial A}{\partial s_e} \right] \cdot 2 \text{Cov}(s_u^2, s_e^2) + \dots \\
& \left. + \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial (s_{ue})^2} + \frac{1}{4} A^{-3/2} \left( \frac{\partial A}{\partial s_{ue}} \right)^2 \right] \text{Var}(s_{ue}) \right\} .
\end{aligned}$$

$$\begin{aligned}
F = & -2 \left\{ \left[ \frac{\partial H}{\partial s_u^2} \frac{\partial G}{\partial s_u^2} - \beta \left( \frac{\partial H}{\partial s_u^2} \right)^2 \right] \text{Var}(s_u^2) + \left[ \frac{\partial H}{\partial s_u^2} \left( \frac{\partial G}{\partial s_e^2} - \beta \frac{\partial H}{\partial s_e^2} \right) \right. \right. \\
& \left. \left. + \frac{\partial H}{\partial s_e^2} \left( \frac{\partial G}{\partial s_u^2} - \beta \frac{\partial H}{\partial s_u^2} \right) \right] \cdot \text{Cov}(s_u^2, s_e^2) + \dots + \left[ \frac{\partial H}{\partial s_{ue}} \left( \frac{\partial G}{\partial s_{ue}} - \beta \frac{\partial H}{\partial s_{ue}} \right) \right] \text{Var}(s_{ue}) \right\} .
\end{aligned}$$

Also

$$N = \tilde{H}^{-2} [J + \tilde{H}E] ,$$

$$\begin{aligned}
E = & (\beta s_u^2 s_{ue}) \left\{ \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial m_{XX}^2} + \frac{1}{4} A^{-3/2} \left( \frac{\partial A}{\partial m_{XX}} \right)^2 \right] \text{Var}(m_{XX}) \right. \\
& + \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial m_{XX} \partial m_{YY}} + \frac{1}{4} A^{-3/2} \frac{\partial A}{\partial m_{XX}} \frac{\partial A}{\partial m_{YY}} \right] 2 \text{Cov}(m_{XX}, m_{YY}) + \dots
\end{aligned}$$

$$\dots + \left[ -\frac{1}{2} A^{-\frac{1}{2}} \frac{\partial^2 A}{\partial m_{XY}^2} + \frac{1}{4} A^{-\frac{3}{2}} \left( \frac{\partial A}{\partial m_{XY}} \right)^2 \right] \text{Var}(m_{XY}) ;$$

and

$$J = -2 \left\{ \left[ \frac{\partial H}{\partial m_{XX}} \left( \frac{\partial G}{\partial m_{XX}} - \beta \frac{\partial H}{\partial m_{XX}} \right) \text{Var}(m_{XX}) + \dots + \left[ \frac{\partial H}{\partial m_{XY}} \left( \frac{\partial G}{\partial m_{XY}} - \beta \frac{\partial H}{\partial m_{XY}} \right) \right] \text{var}(m_{XY}) \right\}.$$

Evaluating each of the above expressions at the probability limits of the corresponding variables, we have

$$D_1 = \frac{1}{d} (-4\bar{a} - 4\bar{A}^{\frac{1}{2}}) \phi ,$$

$$D_2 = \frac{1}{d} (8\bar{a} + 4\bar{A}^{\frac{1}{2}}) \phi ,$$

$$D = -\frac{4}{d} \bar{a}^2 \bar{A}^{-\frac{1}{2}} \phi ,$$

$$F = \frac{16}{d} \bar{a}^3 \phi \bar{A}^{-\frac{1}{2}} \cdot \sigma_x^2 ,$$

Therefore, after some few steps we obtain

$$W + \frac{2(\sigma_u^2 \sigma_e^2 - \sigma_{ue}^2)(\beta \sigma_u^2 - \sigma_{ue})}{(\sigma_x^2)^2 (\beta^2 \sigma_u^2 - 2\beta \sigma_{ue} + \sigma_e^2) \cdot d} \quad (5.20)$$

Also

$$E = \frac{1}{n-1} (-4\bar{a} - 4\bar{a}^2 \bar{A}^{-\frac{1}{2}}) \cdot \phi$$

and

$$J = \frac{1}{n-1} (16\bar{a}^2 \sigma_x^2 + 16\bar{a}^3 \sigma_x^2 \bar{A}^{-\frac{1}{2}}) \cdot \phi .$$

So that, after some algebra we have

$$N = \frac{2(\beta\sigma_u^2 - \sigma_{ue})}{(n-1)(\sigma_x^2)^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2)} [\sigma_x^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2) + (\sigma_u^2\sigma_e^2 - \sigma_{ue}^2)] \quad (5.21)$$

Placing 5.20 and 5.21 in 5.19, we obtain to  $O_p(1/n)$

$$\begin{aligned} E(\hat{\beta} - \beta) &= \frac{\beta\sigma_u^2 - \sigma_{ue}}{(\sigma_x^2)^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2)} \left\{ \frac{1}{n-1} [\sigma_x^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2) \right. \\ &\quad \left. + (\sigma_u^2\sigma_e^2 - \sigma_{ue}^2)] + \frac{1}{d}(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2) \right\} \quad (5.22) \end{aligned}$$

We state our results in the form of a theorem.

Theorem 1 Consider the errors in variables model, one variable case, 5.14. Let  $\hat{\beta}$  given by 5.15 be the maximum likelihood estimator of  $\beta$  in model 5.14. To  $O_p(1/n)$  the asymptotic variance and bias of  $\hat{\beta}$  are given by

$$\bar{E}(\hat{\beta} - \beta)^2 = \frac{1}{(\sigma_x^2)^2} \left\{ \frac{(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2) + \sigma_x^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2)}{n-1} + \frac{(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2)}{d} \right\}, \quad (5.17)$$

and

$$\begin{aligned} \bar{E}(\hat{\beta} - \beta) &= \frac{\beta\sigma_u^2 - \sigma_{ue}}{(\sigma_x^2)^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2)} \left\{ \frac{1}{n-1} [\sigma_x^2(\beta^2\sigma_u^2 - 2\beta\sigma_{ue} + \sigma_e^2) \right. \\ &\quad \left. + (\sigma_u^2\sigma_e^2 - \sigma_{ue}^2)] + \frac{1}{d}(\sigma_u^2\sigma_e^2 - \sigma_{ue}^2) \right\}, \quad (5.22) \end{aligned}$$

respectively.

C. The Asymptotic Bias and Variance of the Maximum Likelihood Estimator. Errors in Variables, Several Variables Case

While the present author was in the process of completing his research on the one variable case, described above, Fuller (1970b) following a quite compact approach obtained the asymptotic expressions for the general case. These expressions are given below for completeness.

Let us write the model as follows:

$$\begin{aligned} Y &= y + e , \\ y &= x\beta , \\ X &= x + u , \end{aligned} \tag{5.23}$$

where  $Y$ ,  $y$  and  $e$  are  $n \times 1$  vectors,  $X$ ,  $x$  and  $u$  are  $n \times k$  matrices.  $(y, x)$  is a matrix of fixed constants. Let

$$A = \begin{bmatrix} \sigma^2 & N' \\ y & H \\ N & H \end{bmatrix} = \bar{M} - \lambda \mathbb{I} ,$$

where  $N$  and  $H$  are submatrices of  $A$ ,  $\bar{M}$  fixed in the limit equals

$$\bar{M} = \begin{bmatrix} \frac{1}{n-1}(y-\bar{y})'(y-\bar{y}) & \frac{1}{n-1}(y-\bar{y})'(x-\bar{x}) \\ \frac{1}{n-1}(x-\bar{x})'(y-\bar{y}) & \frac{1}{n-1}(x-\bar{x})'(x-\bar{x}) \end{bmatrix} .$$

$$\begin{bmatrix} e_j' \\ u_j' \end{bmatrix} \sim \text{independent } N_k(0, \mathbb{I}), j = 1, \dots, n .$$

Assume we have a Wishart matrix  $S$  whose elements are independent

of the errors  $(e_j, u_j)$ , for all  $j$ , such that

$$E(S) = \Sigma .$$

Let us partition  $S$  and  $\Sigma$ , in correspondence to the partition of the variables in 5.23, as follows

$$\Sigma = \begin{bmatrix} s_e^2 & s_{eu} \\ s_{ue} & s_{uu} \end{bmatrix} .$$

Using the above notation we have:

$$A\alpha = 0 ,$$

where  $\alpha' = (1, -\beta')$ . Let us write in addition

$$e - u\beta = v , \text{ and}$$

$$\sigma_v^2 = (1, -\beta') \Sigma \begin{pmatrix} 1 \\ -\beta \end{pmatrix} = \alpha' \Sigma \alpha .$$

Finally, let

$$M = \begin{bmatrix} \frac{1}{n-1} (Y-\bar{Y})'(Y-\bar{Y}) & \frac{1}{n-1} (Y-\bar{Y})'(X-\bar{X}) \\ \frac{1}{n-1} (X-\bar{X})'(Y-\bar{Y}) & \frac{1}{n-1} (X-\bar{X})'(X-\bar{X}) \end{bmatrix}$$

$$= \begin{bmatrix} M_{YY} & M_{YX} \\ M_{XY} & M_{XX} \end{bmatrix} .$$

Using the above notation, the estimator of  $\beta$  is given by

$$[M_{XX} - \lambda_1 s_{uu}] \hat{\beta} = M_{XY} - \lambda_1 s_{ue} , \quad (5.24)$$

$\lambda_1$  being the smallest root of

$$|M - \lambda S| = 0 .$$

We have the following theorem, Fuller (1970b):

Theorem 2 Consider the errors in variables model, several variables case, 5.23. Let  $\hat{\beta}$  given by 5.24 be the maximum likelihood estimator of  $\beta$  in model 5.23. To  $O_p(1/n)$  the asymptotic variance and bias of  $\hat{\beta}$  are given by

$$\bar{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \frac{1}{n} H^{-1} \sigma_v^2 + \left(\frac{1}{n} + \frac{1}{d}\right) H^{-1} (\sigma_v^2 Z_{uv}' - Z_{uv}' Z_{uv}) H^{-1} , \quad (5.25)$$

and

$$\begin{aligned} \bar{E}(\hat{\beta} - \beta) &= -\frac{1}{n} H^{-1} Z_{uv}' - \left(\frac{1}{n} + \frac{1}{d}\right) H^{-1} Z_{uu}' H^{-1} Z_{uv}' \\ &\quad + \frac{1}{\sigma_v^2} \left(\frac{1}{n} + \frac{1}{d}\right) Z_{uv}' H^{-1} Z_{uv}' \cdot H^{-1} Z_{uv}' , \end{aligned} \quad (5.26)$$

where

$$Z_{uv} = Z_{ue} - Z_{uu} \beta$$

and  $d$  are the degrees of freedom associated with  $S$ .

Expressions 5.25 and 5.26 are the ones obtained by Fuller. Together with the author's results, these expressions give a better understanding of the errors in variables problem.

Define  $\hat{H} = M_{XX} - \lambda_1 S_{uu}$  and  $\hat{N} = M_{XY} - \lambda_1 S_{ue}$ . Then an estimator of  $\beta$  unbiased to  $O_p(1/n)$  is

$$\tilde{\beta} = (\hat{H} + \hat{G})^{-1} (\hat{N} + \hat{D}) \quad (5.27)$$

where

$$\hat{G} = \frac{1}{n} \{ S_{uu} + \alpha S_{uu}' \hat{H}^{-1} S_{uu} - \frac{\alpha}{\sigma_v^2} S_{uv}' \hat{H}^{-1} S_{uv} \cdot S_{uu} \} ,$$



$$\hat{D} = \frac{1}{n} \{ S_{ue} + \alpha S_{uu} \hat{H}^{-1} S_{ue} - \frac{\alpha}{\hat{\sigma}_v^2} S_{uv}' \hat{H}^{-1} S_{uv} \cdot S_{ue} \} ,$$

$$\alpha = 1 + \frac{n}{d} ,$$

$$S_{uv} = S_{ue} - S_{uu} \hat{\beta} ,$$

$$\hat{\beta} = \hat{H}^{-1} \hat{N} , \quad \text{and}$$

$$\hat{\sigma}_v^2 = s_e^2 - 2\hat{\beta}' S_{ue} + \hat{\beta}' S_{uu} \hat{\beta} .$$

To  $O_p(1/n)$  the variance of  $\tilde{\beta}$  is the same as that of  $\hat{\beta}$ ; this in turn is given by expression 5.25.

#### D. Comparison of the Maximum Likelihood Estimator with Other Estimators

(i) We first observe that the unbiased maximum likelihood estimator given by 5.27 is more efficient than the estimator studied in Chapter IV. The variance to terms of  $O_p(1/n)$  of the estimator studied in Chapter IV ( $\hat{\beta}$  say), given in the present notation, is equivalent to

$$\bar{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \frac{1}{n} H^{-1} \sigma_v^2 + \left( \frac{1}{n} + \frac{1}{d} \right) H^{-1} (\sigma_v^2 \Sigma_{uu} + \Sigma_{uv}' \Sigma_{uv}) H^{-1} . \quad (5.28)$$

All the matrix terms entering in 5.28 are positive definite. If the trace criterion is used for comparison, it is clear that

$$\text{tr}\{\bar{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\} - \text{tr}\{\bar{E}(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\} = \text{tr}\left\{2\left(\frac{1}{n} + \frac{1}{d}\right) H^{-1} \Sigma_{uu} \Sigma_{uv}' H^{-1}\right\} > 0 .$$

(ii) Consider the one variable case. Assume  $\sigma_e^2$ ,  $\sigma_u^2$  and  $\sigma_{ue}$  are known and that  $\sigma_{ue} = 0$ . We shall consider the following estimator

(Madansky (1959)),

$$\hat{\beta}_1 = \{(m_{YY} - \sigma_e^2)(m_{XX} - \sigma_u^2)^{-1}\}^{\frac{1}{2}} .$$

Using Taylor's expansions it can be shown that to  $O_p(1/n)$

$$\begin{aligned} \bar{E}(\hat{\beta}_1 - \beta^2) &= \frac{1}{4(n-1)(\sigma_x^2)^2} [4\beta^2\sigma_x^2\sigma_u^2 + 2\beta^2(\sigma_u^2)^2 + 4\sigma_x^2\sigma_e^2 + 2\beta^{-2}(\sigma_e^2)^2] \\ &= \frac{1}{(n-1)(\sigma_x^2)^2} \left\{ \sigma_x^2(\beta^2\sigma_u^2 + \sigma_e^2) + \frac{1}{2}[\beta^2(\sigma_u^2)^2 + \beta^{-2}(\sigma_e^2)^2] \right\} . \end{aligned}$$

In particular when  $\sigma_u^2$  and  $\sigma_e^2$  are equal, we have

$$\bar{E}(\hat{\beta}_1 - \beta)^2 = \frac{1}{(n-1)(\sigma_x^2)^2} \left\{ \sigma_x^2(\beta^2\sigma^2 + \sigma^2) + (\sigma^2)^2 \left[ \frac{1}{2}(\beta^2 + \beta^{-2}) \right] \right\} .$$

Now, the function

$$\frac{1}{2}(\beta^2 + \beta^{-2}) \geq 1 , \quad -\infty < \beta < \infty .$$

Therefore,

$$E(\hat{\beta}_1 - \beta)^2 \geq \frac{1}{(n-1)(\sigma_x^2)^2} \{ \sigma_x^2(\beta^2\sigma^2 + \sigma^2) + (\sigma^2)^2 \} = \text{Variance of the M.L.}$$

estimator.

To  $O_p(1/n)$  the bias of  $\hat{\beta}_1$  is given by

$$\bar{E}(\hat{\beta}_1 - \beta) = \frac{1}{4(n-1)(\sigma_x^2)^2} [6\beta\sigma_x^2\sigma_u^2 + 3\beta(\sigma_u^2)^2 - 2\beta^{-1}\sigma_x^2\sigma_e^2 - \beta^{-3}(\sigma_e^2)^2] .$$

(iii) Consider now  $\sigma_u^2 = \sigma_e^2 = \sigma^2$  known and  $\sigma_{ue} = 0$ . Then to  $O_p(1/n)$

$$\hat{\beta}_2 = [m_{XY} / (m_{YY} - \sigma^2)]^{-1}$$

has variance

$$\bar{E}(\hat{\beta}_2 - \beta_2)^2 = \frac{1}{\beta^2 (\sigma_x^2)^2 (n-1)} \{ \sigma_x^2 (\beta^4 \sigma^2 + \beta^2 \sigma^2) + [2(\sigma^2)^2 + \beta^2 (\sigma^2)^2] \} ,$$

which is easily shown to be greater than the variance of the maximum likelihood estimator. Also to  $O_p(1/n)$  the bias of  $\hat{\beta}_2$  is given by

$$\bar{E}(\hat{\beta}_2 - \beta) = \frac{1}{(n-1)(\sigma_x^2)^2} \left[ \beta \sigma_x^2 \sigma^2 + \frac{(\sigma^2)^2}{\beta} \right] .$$

# VI. ESTIMATION IN A GENERALIZED COVARIANCE MODEL WITH ERRORS IN THE COVARIATES

A covariance model with errors in the covariates has been studied by several authors. DeGracie (1968) gives references on the problem and proceeds to develop estimators unbiased to  $O(1/r)$  ( $r$  = No. of replications). An extension of his work is given below in terms of generalized inverses.

## A. The Covariance Model with Errors in the Covariates

Let us consider the model

$$\begin{aligned} Y &= X\beta + z\gamma + e, \\ Z &= z + v, \end{aligned} \quad (6.1)$$

where the  $n \times q$  matrix  $X$  of fixed constants is of rank  $r \leq q$ ,  $Y$  is a  $n \times 1$  vector of observations,  $Z$  and  $z$  are  $n \times p$  matrices,  $v$  a  $n \times p$  matrix of errors and  $e$  a  $n \times 1$  vector of errors. Let  $(v_i, e_i)$  be the  $i$ th row of the matrix  $(v, e)$ . The joint distribution of  $(v_i, e_i)'$  is  $(p+1)$  - variate normal with

$$E(v_i, e_i)' = 0 \quad 1 \leq i \leq n$$

and

$$E(v_i, e_i)'(v_i, e_i) = \begin{bmatrix} \phi & \psi \\ \psi' & \sigma_e^2 \end{bmatrix}, \quad 1 \leq i \leq n.$$

$(v_i, e_i)'$  distributed independently of  $(v_j, e_j)'$  if  $i \neq j$ ,  $1 \leq i, j \leq n$ . We assume that all columns of  $Z$  are linearly independent of the columns of  $X$  and the rank of  $Z$  is equal to  $p$ . In a sense the model defined by expressions 6.1 is a special case of the general errors in variables

model studied in a previous chapter. However, model 6.1 is also more general than the errors in variables model in that we permit the rank of  $X$  to be less than  $q$ .

Consider the expression

$$\begin{bmatrix} \frac{1}{n}X'X & \frac{1}{n}X'Z \\ \frac{1}{n}Z'X & \frac{1}{n}Z'Z-S^* \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}X'Y \\ \frac{1}{n}Z'Y-R^* \end{bmatrix},$$

where  $S^*$  and  $R^*$  are estimators such that  $E(S^*) = \phi$  and  $E(R^*) = \psi$ .

Then,

$$E(X'Z) = E[X'(z+v)] = X'z$$

$$E(Z'Z) = E[(z+v)'(z+v)] = z'z + n\phi.$$

Now let us examine

$$E(X'Y) = E[X'(X\beta + z\gamma + e)] = X'X\beta + X'z\gamma$$

and

$$E(Z'Y) = E[(z+v)'(X\beta + z\gamma + e)] = z'X\beta + z'z\gamma + n\psi.$$

Using the above results, we find

$$E \begin{bmatrix} \frac{1}{n}X'X & \frac{1}{n}X'Z \\ \frac{1}{n}Z'X & \frac{1}{n}Z'Z-S^* \end{bmatrix} = \begin{bmatrix} \frac{1}{n}X'X & \frac{1}{n}X'z \\ \frac{1}{n}z'X & \frac{1}{n}z'z \end{bmatrix}$$

and

$$E \begin{bmatrix} \frac{1}{n}X'X & \frac{1}{n}X'Z \\ \frac{1}{n}Z'X & \frac{1}{n}Z'Z-R^* \end{bmatrix} = \begin{bmatrix} \frac{1}{n}X'X & \frac{1}{n}X'z \\ \frac{1}{n}z'X & \frac{1}{n}z'z \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} .$$

We can see that the right hand sides of the above last two equations look very much like  $H$  and  $N$  of Chapter IV. The only difference between the two cases is the starting model. In the present case we are dealing with a non-full rank model. We shall show how the methods developed for the case where  $H$  (and consequently  $\hat{H}$ ) is nonsingular can also be used when  $H$  (or  $\hat{H}$ ) is singular.

Since the matrix  $X'X$  is singular, we would not be able to estimate all of the  $\beta$  parameters. However, our objective when working with this kind of model is an estimator for any relevant contrast. Let  $W$  be a  $n \times r$  matrix whose column space (the space generated by its columns) is identical with the column space of the matrix  $X$  of model 6.1. Because the rank of  $X$  equals  $r$ , all the columns of the matrix  $W$  will be linearly independent. Then we know that corresponding to every  $n \times 1$  vector  $\alpha$  belonging to the column space of  $X$  there exists a  $q \times 1$  vector  $\beta$  and a  $r \times 1$  vector  $\theta$  such that

$$\alpha = X\beta = W\theta .$$

Thus, an alternative expression for model 6.1 is

$$Y = W\theta + z\gamma + e$$

$$Z = z + v . \quad (6.2)$$

Model 6.2 is a nonsingular "reparametrization" of model 6.1. The whole

methodology of estimation developed for the errors in variables model introduced in Chapter IV, can now be used with the model written in the form

6.2. Using form 6.2 we arrive at the system

$$\begin{bmatrix} \frac{1}{n}W'W & \frac{1}{n}W'Z \\ \frac{1}{n}Z'W & \frac{1}{n}Z'Z-S^* \end{bmatrix} \begin{bmatrix} \hat{\theta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}W'Y \\ \frac{1}{n}Z'Y-R^* \end{bmatrix}, \quad (6.3)$$

where the matrix

$$\begin{bmatrix} \frac{1}{n}W'W & \frac{1}{n}W'Z \\ \frac{1}{n}Z'W & \frac{1}{n}Z'Z-S^* \end{bmatrix}$$

is nonsingular. Placing the values of  $Y$  and  $Z$  given by equations 6.1, in 6.2, we obtain

$$\begin{aligned} & \left[ \begin{bmatrix} \frac{1}{n}W'W & \frac{1}{n}W'Z \\ \frac{1}{n}Z'W & \frac{1}{n}Z'Z \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{n}W'v \\ \frac{1}{n}v'W & \frac{1}{n}(z'v+v'z+v'v)-S^* \end{bmatrix} \right] \begin{bmatrix} \hat{\theta} \\ \hat{\gamma} \end{bmatrix} \\ &= \left[ \begin{bmatrix} \frac{1}{n}W'y \\ \frac{1}{n}Z'y \end{bmatrix} + \begin{bmatrix} \frac{1}{n}W'e \\ \frac{1}{n}(v'y+v'e+z'e) -R^* \end{bmatrix} \right], \quad (6.4) \end{aligned}$$

where we wrote  $y = X\beta + z\gamma$ . Let us write

$$H = \begin{bmatrix} \frac{1}{n} W'W & \frac{1}{n} W'Z \\ \frac{1}{n} Z'W & \frac{1}{n} Z'Z \end{bmatrix},$$

$$a = \begin{bmatrix} 0 & \frac{1}{n} W'v \\ \frac{1}{n} v'W & \frac{1}{n} (Z'v + v'Z + v'v) - S^* \end{bmatrix},$$

$$N = \begin{bmatrix} \frac{1}{n} W'u \\ \frac{1}{n} Z'y \end{bmatrix}$$

and

$$b = \begin{bmatrix} \frac{1}{n} W'e \\ \frac{1}{n} (v'y + v'e + Z'e) - R^* \end{bmatrix},$$

Also let

$$\hat{H} = \begin{bmatrix} \frac{1}{n} W'W & \frac{1}{n} W'Z \\ \frac{1}{n} Z'W & \frac{1}{n} Z'Z - S^* \end{bmatrix} \quad (6.5)$$



and

$$\hat{N} = \begin{bmatrix} \frac{1}{n} W'Y \\ \frac{1}{n} Z'Y - R^* \end{bmatrix} . \quad (6.6)$$

The matrix  $(u, e)$  of errors from Chapter IV is now of the form  $(0, v, e)$  with

$$E(u_i, e_i)' = 0 , \quad 1 \leq i \leq n$$

and

$$E(u_i, e_i)'(u_i, e_i) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \emptyset & \psi \\ 0 & \psi' & \sigma_e^2 \end{bmatrix} \quad 1 \leq i \leq n ;$$

$(u_i, e_i)'$  distributed independently of  $(u_j, e_j)'$  if  $i \neq j$ ,  $1 \leq i, j \leq n$ .

We have in addition

$$S = \begin{bmatrix} 0 & 0 \\ 0 & S^* \end{bmatrix} \quad (6.7)$$

and

$$R = \begin{bmatrix} 0 \\ R^* \end{bmatrix} ; \quad (6.8)$$

to match the notation used in Chapter IV. Also, in 6.7 and 6.8  $E(S^*) = \emptyset$  and  $E(R^*) = \psi$ , respectively.

Our covariance model can now be considered as a special case of the general errors in variables model. Fuller's (1968) method of estimation can now be used. We shall show that for our problem the matrices  $\hat{G}$  and  $\hat{D}$  of Chapter IV take a special form. Let us write  $\hat{H}^{-1}$ , where  $\hat{H}$  is defined by 6.5, as follows.

$$\hat{H}^{-1} = \begin{bmatrix} \hat{H}^{11} & \hat{H}^{12} \\ \hat{H}^{21} & \hat{H}^{22} \end{bmatrix}.$$

Then

$$\hat{H}^{-1}S = \begin{bmatrix} \hat{H}^{11} & \hat{H}^{12} \\ \hat{H}^{21} & \hat{H}^{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & S^* \end{bmatrix} = \begin{bmatrix} 0 & \hat{H}^{12}S^* \\ 0 & \hat{H}^{22}S^* \end{bmatrix}$$

therefore

$$\text{tr}(\hat{H}^{-1}S) = \text{tr}(\hat{H}^{22}S^*) \quad (6.9)$$

Also

$$S\hat{H}^{-1}S = \begin{bmatrix} 0 & 0 \\ 0 & S^* \end{bmatrix} \begin{bmatrix} \hat{H}^{11} & \hat{H}^{12} \\ \hat{H}^{21} & \hat{H}^{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & S^* \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & S^*\hat{H}^{22}S^* \end{bmatrix} \quad (6.10)$$

Using 6.7, 6.8, 6.9 and 6.10 we see that  $\hat{G}$  and  $\hat{D}$  as defined in Chapter IV are of the form

$$\hat{G} = \begin{bmatrix} 0 & 0 \\ 0 & \hat{G}^* \end{bmatrix},$$

$$\hat{D} = \begin{bmatrix} 0 \\ \hat{D}^* \end{bmatrix}.$$

$\hat{G}^*$  and  $\hat{D}^*$  in the above expressions are given by

$$\hat{G}^* = \frac{1}{n}[k+1+\alpha \text{tr}(\hat{H}^{22}S^*)]S^* + \left(\frac{1}{n} + \frac{1}{d}\right)S^*\hat{H}^{22}S^* \quad (6.11)$$

and

$$\hat{D}^* = \frac{1}{n}[k+1+\alpha \text{tr}(\hat{H}^{22}S^*)]R^* + \left(\frac{1}{n} + \frac{1}{d}\right)S^*\hat{H}^{22}R^* \quad (6.12)$$

where  $k = p+r$ , the rank of  $\hat{H}$  and  $\alpha = 1 + n/d$ ,  $d$  being the degrees of freedom associated with  $S^*$ . Our nearly unbiased estimators are obtained by solving the system

$$\left[ \begin{pmatrix} \frac{1}{n}W'W & \frac{1}{n}W'Z \\ \frac{1}{n}Z'W & \frac{1}{n}Z'Z-S^* \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \hat{G}^* \end{pmatrix} \right] \begin{bmatrix} \theta \\ \gamma \end{bmatrix} = \left[ \begin{pmatrix} \frac{1}{n}W'Y \\ \frac{1}{n}Z'Y-R^* \end{pmatrix} + \begin{pmatrix} 0 \\ \hat{D}^* \end{pmatrix} \right] . \quad (6.13)$$

Let us rewrite system 6.13 as follows:

$$\begin{aligned} \frac{1}{n}W'W\theta + \frac{1}{n}W'Z\gamma &= \frac{1}{n}W'Y \\ \frac{1}{n}Z'W\theta + \frac{1}{n}Z'Z\gamma - S^*\gamma + \hat{G}^*\gamma &= \frac{1}{n}Z'Y - R^* + \hat{D}^* \end{aligned} \quad (6.14)$$

If we multiply the first of the above expressions on the left by  $Z'W(W'W)^{-1}$  we obtain

$$Z'W(W'W)^{-1}W'W\theta + Z'W(W'W)^{-1}W'Z\gamma = Z'W(W'W)^{-1}W'Y \quad (6.15)$$

Now observe that  $W(W'W)^{-1}W' = X(X'X)^*X'$  the orthogonal projection operator on the column space of the matrix  $X$ , where the notation  $(X'X)^*$  is used to denote the conditional inverse of  $X'X$ . Since corresponding to  $\theta$  there is a vector  $\beta$  such that  $W\theta = X\beta$ , we can write 6.15 in terms of the original variables in model 6.1, as follows

$$Z'X(X'X)^*X'X\beta + Z'X(X'X)^*X'Z\gamma = Z'X(X'X)^*X'Y \quad (6.16)$$

The second expression 6.14 becomes

$$\frac{1}{n}Z'X\beta + \frac{1}{n}Z'Z\gamma - S^*\gamma + \hat{G}^*\gamma = \frac{1}{n}Z'Y - R^* + \hat{D}^* \quad (6.17)$$

Let us write  $M = X(X'X)^*X'$ , then we know that  $MX = X$  so that from 6.16 we obtain

$$\frac{1}{n}Z'X\beta = \frac{1}{n}Z'MY - \frac{1}{n}Z'MZ\gamma \quad .$$

Substituting the above value in 6.17, we obtain after some rearrangement

$$\left[\frac{1}{n}Z'(I-M)Z-S^* + \hat{G}^*\right]\gamma = \frac{1}{n}Z'(I-M)Y - R + \hat{D}^* \quad (6.18)$$

Multiplying the first expression 6.14 on the left by  $W(W'W)^{-1}$ , we obtain

$$\frac{1}{n}W(W'W)^{-1}W'W\theta + \frac{1}{n}W(W'W)^{-1}W'Z\beta = \frac{1}{n}W(W'W)^{-1}W'Y$$

which in turn can be written

$$\begin{aligned} \frac{1}{n}X(X'X)^{-1}X'X\beta + \frac{1}{n}X(X'X)^{-1}X'Z\gamma &= \frac{1}{n}X(X'X)^{-1}X'Y \\ \frac{1}{n}X\beta + \frac{1}{n}MX\gamma &= \frac{1}{n}MY \end{aligned} \quad (6.19)$$

Multiplying 6.19 on the left by  $X'$ , we obtain

$$\frac{1}{n}X'X\beta + \frac{1}{n}X'Z\gamma = \frac{1}{n}X'Y \quad (6.20)$$

Expressions 6.18 and 6.20 together are equivalent to expressions 6.14.

Thus, we have shown the following theorem.

Theorem 1      There always exists a nonsingular reparametrization of model 6.1. The system of "normal equations" to obtain estimators of any relevant contrast is given by expressions 6.21:

$$\begin{aligned} \frac{1}{n}X'X\beta + \frac{1}{n}X'Z\gamma &= \frac{1}{n}X'Y \\ \left[\frac{1}{n}Z'(I-M)Z-S^* + \hat{G}^*\right]\gamma &= \frac{1}{n}Z'(I-M)Y - R^* + \hat{D}^* \end{aligned} \quad (6.21)$$

where  $\hat{G}^*$  and  $\hat{D}^*$  are given by expressions 6.11 and 6.12 and  $M = X(X'X)^{-1}X'$ .

The estimators obtained from system 6.21 are unbiased to  $O_p(1/n)$ .

Now let us consider the estimation of any linear parametric function  $v'\beta$  where  $v'$  can be expressed as a linear combination of the rows of  $X$ .

That is, a parametric function for which  $v' = a'X$  for some  $n \times 1$  vector  $a$ . Let  $\hat{v}'\beta = a'X\hat{\beta}$  be the desired estimator and  $\hat{\gamma}$  the solution of 6.21, then

$$\hat{v}'\beta = a'X\hat{\beta} = a'M(Y - Z\hat{\gamma})$$

where we have used the result given by 6.20. Restricting the above class of estimators to those for which  $a = X\rho$  or  $v' = \rho'X'X$  for some  $q \times 1$  vector  $\rho$ , we obtain

$$\hat{v}'\beta = \rho'X'M(Y - Z\hat{\gamma}) = \rho'X'Y - \rho'X'Z\hat{\gamma}. \quad (6.22)$$

Observing the similarities between the above approach and the theory of estimation under the general linear hypothesis model, we call the system of equations

$$\left[ \frac{1}{n}Z'(I-M)Z - S^* + \hat{G}^* \right] \hat{\gamma} = \frac{1}{n}Z'(I-M)Y - R^* + \hat{D}^*$$

"the reduced system for estimating  $\gamma$  eliminating  $\beta$ ". Therefore, knowledge of the estimator of  $\beta$  in the reduced model

$$Y = X\beta + e$$

can be exploited for purposes of estimation under the full model 6.1. Also, if the above model is the general linear hypothesis model, then  $\rho'X'Y$  would be the best linear unbiased estimator of the "estimable" linear parametric function  $v'\beta$ .

An additional observation will help to simplify the application of the above methodology. Consider the matrix  $\hat{H}^{22}$  which appears in  $\hat{G}^*$  and  $\hat{D}^*$ . From matrix theory, and using our previous notation:

$$\hat{H}^{22} = [\hat{H}_{22} - \hat{H}_{21}\hat{H}_{11}^{-1}\hat{H}_{12}]^{-1}$$

$$= \left[ \frac{1}{n} Z' Z - S^* - \frac{1}{n} Z' W (W' W)^{-1} W' Z \right]^{-1}$$

$$\hat{H}^{22} = \left[ \frac{1}{n} Z' (I - M) Z - S^* \right]^{-1} \quad (6.23)$$

The estimation procedure for a covariance model with errors in the covariates, can be summarized as follows:

(i) Consider the reduced model  $Y = X\beta + e$  as if we had the general linear hypothesis model and obtain the necessary knowledge about the matrix  $(X'X)^*$ . i.e., obtain a particular solution to the system  $X'X\beta = X'Y$  for all  $Y$ 's.

(ii) Construct "the reduced system of equations for estimating  $\gamma$  eliminating  $\beta$ ". i.e., construct the system

$$\left\{ \frac{1}{n} Z' (I - M) Z - S^* + \hat{G}^* \right\} \hat{\gamma} = \left\{ \frac{1}{n} Z' (I - M) Y - R^* + \hat{D}^* \right\},$$

where  $\hat{G}^*$  and  $\hat{D}^*$  are given by expressions 6.11 and 6.12, respectively, and solve for  $\hat{\gamma}$ . The matrix inside the brackets on the left hand side of the above system will in general be nonsingular.

(iii) The estimator of any linear parametric function  $v'\beta$  for which  $v' = \rho'X'X$ , for some vector  $\rho$ , is then given by

$$v\hat{\beta} = \rho'X'[Y - Z\hat{\gamma}].$$

(iv) To  $O_p(1/n)$

$$E(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)' = \frac{1}{n} H^{22} \sigma_v^2 + \left( \frac{1}{n} + \frac{1}{d} \right) \sigma_v^2 H^{22} \phi H^{22} + \left( \frac{1}{n} + \frac{1}{d} \right) H^{22} (\psi - \phi\gamma)(\psi - \phi\gamma)' H^{22},$$

where  $H^{22} = [Z'(I - M)Z]^{-1}$  and  $\sigma_v^2 = \sigma_e^2 - 2\gamma'\psi + \gamma'\phi\gamma$ .

### B. The Maximum Likelihood Estimator

We observe that the maximum likelihood method could also be used if we had available an unbiased estimator of  $E(v_i, e_i)'(v_i, e_i)$  of Section A. The estimator is to be distributed independent of  $(v_i, e_i)$  for all  $i$ .

Let

$$\begin{bmatrix} \hat{\phi} & \hat{\psi} \\ \hat{\psi}' & \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} S_{uu} & S_{ue} \\ S_{eu} & s_e^2 \end{bmatrix} \quad (6.24)$$

be such an estimator. Reasoning in the same fashion as in Section A above, we would arrive at the system of normal equations

$$X'X\beta + X'Z\gamma = X'Y$$

$$\left[ \frac{1}{n} Z'(I-M)Z - \lambda_1 S_{uu} \right] \gamma = \frac{1}{n} Z'(I-M)Y - \lambda_1 S_{ue} \quad ;$$

where  $\lambda_1$  is the minimum root of the determinantal equation

$$\left| \begin{bmatrix} \frac{1}{n} Z'(I-M)Z & \frac{1}{n} Z'(I-M)Y \\ \frac{1}{n} Y'(I-M)Z & \frac{1}{n} Y'(I-M)Y \end{bmatrix} - \lambda_1 \begin{bmatrix} S_{uu} & S_{ue} \\ S_{eu} & s_e^2 \end{bmatrix} \right| = 0 \quad .$$

Applying the methods of Chapter III and the results of Chapter IV, define

$$\hat{G} = \frac{1}{n} \{ S_{uu} + \alpha S_{uu} [Z'(I-M)Z]^{-1} S_{uu} - \frac{\alpha}{\hat{\sigma}_e^2} S'_{uv} [Z'(I-M)Z]^{-1} S_{uv} \cdot S_{uu} \}$$

and

$$\hat{D} = \frac{1}{n} \{ S_{ue} + \alpha S_{uu} [Z'(I-M)Z]^{-1} S_{ue} - \frac{\alpha}{\hat{\sigma}_e^2} S'_{uv} [Z'(I-M)Z]^{-1} S_{uv} \cdot S_{ue} \} \quad ,$$

where  $d$  are the degrees of freedom associated with 6.24 and

$$\alpha = 1 + \frac{n}{d} ,$$

$$S_{uv} = S_{ue} - S_{uu} \hat{\gamma} ,$$

$$\hat{\gamma} = \left[ \frac{1}{n} Z' (I-M) Z - \lambda_1 S_{uu} \right]^{-1} \left[ \frac{1}{n} Z' (I-M) Y - \lambda_1 S_{ue} \right] ,$$

and

$$\sigma_v^2 = s_e^2 - 2 \hat{\gamma}' S_{ue} + \hat{\gamma}' S_{uu} \hat{\gamma} .$$

Then an estimator for  $\gamma$  unbiased to  $O_p(1/n)$  is given by

$$\tilde{\gamma} = \left[ \frac{1}{n} Z' (I-M) Z - \lambda_1 S_{uu} + \hat{G} \right]^{-1} \left[ \frac{1}{n} Z' (I-M) Y - \lambda_1 S_{ue} + \hat{D} \right] . \quad (6.25)$$

We observe that under very special circumstances we would have an unbiased estimator of  $E(v_i, e_i)'(v_i, e_i)$  independent of  $(v_i, e_i)$  for all  $i = 1, 2, \dots, n$ , with the Wishart distribution. This will be the case if we have orthogonal repetitions of a basic experiment.

Also, to  $O_p(1/n)$

$$\begin{aligned} E(\tilde{\gamma} - \gamma)(\tilde{\gamma} - \gamma)' &= \frac{1}{n} H^{22} \sigma_v^2 + \left( \frac{1}{n} + \frac{1}{d} \right) H^{22} \phi H^{22} \sigma_v^2 \\ &\quad - \left( \frac{1}{n} + \frac{1}{d} \right) H^{22} (\psi - \phi \gamma)(\psi - \phi \gamma)' H^{22} , \end{aligned} \quad (6.26)$$

where  $H^{22} = [Z'(I-M)Z]^{-1}$  and  $\sigma_v^2 = \sigma_e^2 - 2\gamma'\psi + \gamma'\phi\gamma$ .  $\tilde{\gamma}$ , the maximum likelihood estimator given by 6.25, is clearly more efficient than the modified estimator given by 6.21 (to  $O_p(1/n)$ ).



### C. The Maximum Likelihood Estimator for a Simultaneous Equations Problem

The simultaneous equations problem studied in Section C of Chapter IV, can also be studied using the maximum likelihood approach. Because of invariance of sums of squares and products under orthogonal transformations, we can work with the original variables. Let us write the errors in variables form of the simultaneous equations problem as follows

$$y_1 = -X_1\gamma_1' - \hat{Y}_2\beta_2' + e_1 ,$$

where

$$\hat{Y}_2 = X(X'X)^{-1}X'Y_2 .$$

The notation is now that introduced in Chapter IV. Then we know that

$$E\left\{\frac{(Y_2, Y_1)'(I-M)(Y_2, Y_1)}{n-K}\right\} = 0$$

where  $M = X(X'X)^{-1}X'$ . Let  $M_1 = X_1(X_1'X_1)^{-1}X_1'$  then, we arrive at the system of normal equations

$$X_1'X_1(-\gamma_1') + X_1'\hat{Y}_2(-\beta_2') = X_1'y_1 ,$$

$$\left[\frac{1}{n} Y_2'(M-M_1)Y_2 - \lambda_1 KS^*\right](-\beta_2') = \frac{1}{n} Y_2'(M-M_1)y_1 - \lambda_1 KR^* ,$$

where  $\lambda_1$  is the minimum root of the determinantal equation

$$\left| \begin{bmatrix} Y_2' \\ y_1' \end{bmatrix} (M-M_1)(Y_2, y_1) - \lambda_1 K \begin{bmatrix} Y_2' \\ y_1' \end{bmatrix} (I-M)(Y_2, y_1) \right| = 0 ,$$

and

$$\frac{1}{n} \hat{\phi} = \begin{bmatrix} S^* & R^* \\ R^{*'} & s_e^2 \end{bmatrix} = \frac{(Y_2, y_1)' (I-M)(Y_2, y_1)}{n(n-K)} .$$

The elements of the last matrix are all  $O_p(1/n)$ . Let

$$\begin{aligned} G^* &= \{S^* + \alpha S^* [Y_2' - (I-M_1)Y_2]^{-1} S^* - \frac{\alpha}{\hat{\sigma}_{uv}^2} S^{*'} [Y_2' (I-M_1)Y_2]^{-1} S_{uv}^* \cdot S^*\} ; \\ D^* &= \{R^* + \alpha S^* [Y_2' - (I-M_1)Y_2]^{-1} R^* - \frac{\alpha}{\hat{\sigma}_{uv}^2} S^{*'} [Y_2' (I-M_1)Y_2]^{-1} S_{uv}^* \cdot R^*\} \quad (6.27) \end{aligned}$$

where  $\alpha = (1 + \frac{n}{d})$  and

$$\begin{aligned} S_{uv}^* &= R^* - S^* (-\hat{\beta}_2') , \\ -\hat{\beta}_2' &= [\frac{1}{n} Y_2' (M-M_1)Y_2 - \lambda_1 K S^*]^{-1} [\frac{1}{n} Y_2' (M-M_1)y_1 - \lambda_1 K R^*] , \\ \hat{\sigma}_{uv}^2 &= s_e^2 - 2 \hat{\beta}_2' R^* + \hat{\beta}_2' S^* \hat{\beta}_2' . \end{aligned}$$

Then, corresponding to 6.25 an estimator for  $\beta_2$  unbiased to  $O_p(1/n)$  is obtained by solving the system

$$[\frac{1}{n} Y_2' (M-M_1)Y_2 - \lambda_1 K S^* + G^*] (-\beta_2') = [\frac{1}{n} Y_2' (M-M_1)y_1 - \lambda_1 K R^* + D^*] . \quad (6.28)$$

We observe in addition that all terms entering in  $G^*$  are  $O_p(1/n^2)$  except  $S^*$  which is  $O_p(1/n)$ . Similarly, all terms entering in  $D^*$  are  $O_p(1/n^2)$  except  $R^*$  which is  $O_p(1/n^2)$ . We can neglect all terms of  $O_p(1/n^2)$  or smaller and take  $G^* = S^*$ , and  $D^* = R^*$ .

The estimator of  $\beta'_2$  given by 6.28 has to  $O_p(1/n)$  the same asymptotic variance than the estimator derived in Chapter IV.

#### D. The Randomized Complete Block Design with Two Covariates Subject to Error

Let us consider a randomized complete block design where on each experimental unit we observe two covariates. We shall discuss first the estimation procedure when both covariates are measured without error. The model is

$$Y_{ij} = \mu + b_i + t_j + \gamma_1 Z_{ij}^{(1)} + \gamma_2 Z_{ij}^{(2)} + e_{ij} ; \quad (6.29)$$

$$i = 1, \dots, r, \quad j = 1, \dots, t;$$

where the errors are uncorrelated with mean zero and constant variance  $\sigma^2$ .

We shall work first with the classification part of the model; i.e., the model

$$Y_{ij} = \mu + b_i + t_j + e_{ij} \quad i = 1, \dots, r, \quad j = 1, \dots, t.$$

Imposing the restrictions  $\sum t_j = 0$  and  $\sum b_i = 0$  we obtain (using the dot notation)

$$\hat{t}_j = \frac{1}{r} Y_{.j} - \frac{1}{rt} Y_{..} , \quad j = 1, 2, \dots, t ,$$

$$\hat{b}_i = \frac{1}{t} Y_{i.} - \frac{1}{rt} Y_{..} , \quad i = 1, 2, \dots, r ,$$

Under the above conditions, the error sum of squares whose matrix expression is  $Y'(I-M)Y$ , is given by

$$Y'(I-M)Y = \sum_{ij} Y_{ij}^2 - \sum_i \frac{Y_{i.}^2}{t} - \sum_j \frac{Y_{.j}^2}{r} + \frac{1}{rt} Y_{..}^2 . \quad (6.30)$$

The estimation problem for model 6.29 will be completely solved if we exploit the solution for the classification part of the model, to construct the reduced normal equations for  $\gamma_1$  and  $\gamma_2$ . That is, following Zyskind (1964) to construct the system

$$Z'(I-M)ZY = Z'(I-M)Y, \quad (6.31)$$

where  $\gamma' = (\gamma_1, \gamma_2)$ . System 6.31 is easily constructed substituting the columns of  $Z$  for  $Y$  in 6.30. For example, if  $E_Z(i)_Z(j)$  is the element on the  $i$ th row,  $j$ th column of  $Z'(I-M)Z$ , we can write

$$E_Z(1)_Z(1) = \sum_{i,j} (Z_{ij}^{(1)})^2 - \sum_i \frac{(Z_{i.}^{(1)})^2}{t} - \sum_j \frac{(Z_{.j}^{(1)})^2}{r} + \frac{(Z_{..}^{(1)})^2}{rt},$$

$$E_Z(1)_Z(2) = \sum_{i,j} (Z_{ij}^{(1)} Z_{ij}^{(2)}) - \sum_i \frac{Z_{i.}^{(1)} Z_{i.}^{(2)}}{t} - \sum_j \frac{Z_{.j}^{(1)} Z_{.j}^{(2)}}{r} + \frac{Z_{..}^{(1)} Z_{..}^{(2)}}{rt},$$

$$E_Z(2)_Z(2) = \sum_{i,j} (Z_{ij}^{(2)})^2 - \sum_i \frac{(Z_{i.}^{(2)})^2}{t} - \sum_j \frac{(Z_{.j}^{(2)})^2}{r} + \frac{(Z_{..}^{(2)})^2}{rt}.$$

To construct the left hand side of 6.31 we proceed as follows. Let

$E_Z(i)_Y$  be the  $i$ th element of the vector  $Z'(I-M)Y$ . Then

$$E_Z(1)_Y = \sum_{i,j} Z_{ij}^{(1)} Y_{ij} - \sum_i \frac{Z_{i.}^{(1)} Y_{i.}}{t} - \sum_j \frac{Z_{.j}^{(1)} Y_{.j}}{r} + \frac{Z_{..}^{(1)} Y_{..}}{rt}$$

and

$$E_Z(2)_Y = \sum_{i,j} Z_{ij}^{(2)} Y_{ij} - \sum_i \frac{Z_{i.}^{(2)} Y_{i.}}{t} - \sum_j \frac{Z_{.j}^{(2)} Y_{.j}}{r} + \frac{Z_{..}^{(2)} Y_{..}}{rt}.$$

Then 6.31 takes the form

$$\begin{bmatrix} E_{Z(1)Z(1)} & E_{Z(1)Z(2)} \\ E_{Z(2)Z(1)} & E_{Z(2)Z(2)} \end{bmatrix} \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} E_{Z(1)Y} \\ E_{Z(2)Y} \end{bmatrix} \quad (6.32)$$

The least squares estimators  $\gamma_1$  and  $\gamma_2$  are therefore obtained by solving 6.32 for  $(\hat{\gamma}_1, \hat{\gamma}_2)'$ . The matrix of E's operating on the vector of  $\hat{\gamma}$ 's is in this case nonsingular.

We shall now apply the methodology developed in previous sections when the covariates are subject to error. We shall assume the matrix  $\phi$  to be known (for which case it is sufficient to let  $d = +\infty$ ) and  $E(v'e) = \psi = 0$ . Let us write

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix},$$

then from expression 6.11, we have

$$\begin{aligned} \hat{G}^* &= \frac{1}{n} [r+3 + \text{tr}(\left(\frac{1}{n} Z'(I-M)Z - \phi\right)^{-1} \cdot \phi)] \cdot \phi \\ &= \frac{1}{n} \left\{ r+3 + \text{tr} \left( \left[ \frac{1}{n} \begin{bmatrix} E_{Z(1)Z(1)} & E_{Z(1)Z(2)} \\ E_{Z(2)Z(1)} & E_{Z(2)Z(2)} \end{bmatrix} - \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \right]^{-1} \right. \right. \\ &\quad \left. \left. \cdot \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \right) \right\} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}. \end{aligned}$$

Since  $\psi = 0$ , then  $R^* = 0$ .

Observe that  $p = \text{rank of } H^{22} = 2 = \text{number of covariates}$ . The above value of  $G^*$  is placed in 6.21 to obtain

$$\frac{1}{n} \begin{bmatrix} E_{Z(1)Z(1)} & E_{Z(1)Z(2)} \\ E_{Z(2)Z(1)} & E_{Z(2)Z(2)} \end{bmatrix} - \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} + \hat{G}^* \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} E_{Z(1)Y} \\ E_{Z(2)Y} \end{bmatrix}.$$

Estimators for  $\gamma_1$  and  $\gamma_2$  unbiased to  $O_p(1/n)$  are obtained by solving the system for  $(\hat{\gamma}_1, \hat{\gamma}_2)'$ . By imposing the restrictions  $\sum_i b_i = 0$  and  $\sum_j t_j = 0$  we shall obtain the following results

$$\begin{aligned} \hat{u} &= \frac{Y_{..}}{rt} - \hat{\gamma}_1 \frac{X_{..}^{(1)}}{rt} - \hat{\gamma}_2 \frac{X_{..}^{(2)}}{rt}, \\ \hat{b}_i &= \frac{Y_{i.}}{t} - \frac{Y_{..}}{rt} - \hat{\gamma}_1 \frac{X_{i.}^{(1)}}{t} - \hat{\gamma}_2 \frac{X_{i.}^{(2)}}{t}, \quad i = 1, \dots, r \\ t_j &= \frac{Y_{.j}}{r} - \frac{Y_{..}}{rt} - \hat{\gamma}_1 \frac{X_{.j}^{(1)}}{r} - \hat{\gamma}_2 \frac{X_{.j}^{(2)}}{r}, \quad j = 1, \dots, t. \end{aligned}$$

The above estimators are obtained by solving the corresponding normal equations replacing  $\gamma_1$  and  $\gamma_2$  by their estimators.

## VII. A SAMPLING EXPERIMENT

In this chapter we describe the results of an empirical investigation in order to observe the behavior of four estimators. Namely:

- (i) the two stage least squares estimator modified to have finite variance;
- (ii) the limited information maximum likelihood estimator modified to have finite variance;
- (iii) the limited information maximum likelihood estimator unbiased to  $O_p(1/n)$ , and
- (iv) a second modification of the two stage least squares estimator possessing finite variance.

We present the basic theory of the method, the sampling procedure and the results of Monte Carlo experiment.

## A. The Basic Theory

Consider the first equation

$$y_1 = Y_2\beta + X_1\gamma + u_1 \quad (7.1)$$

of a system of simultaneous equations, where

$y_1$  and  $u_1$  are  $n \times 1$  vectors ,

$Y_2$  is  $n \times (G_1 - 1)$  ,

$X_1$  is  $n \times K_1$  ,

$\beta$  is  $G_1 - 1$  , and

$\gamma$  is  $K_1 \times 1$  .

The reduced form for the endogenous variables in the equation is written

$$(y_{1t}, y_{2t}) = (x_{1t}, x_{2t}) \begin{bmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{bmatrix} + (v_{1t}, v_{2t}), \quad t = 1, \dots, n, \quad (7.2)$$

or

$$y_t = x_t \pi + v_t.$$

The dimensions in 7.2 are as follows

$y_{1t}$  and  $v_{1t}$  are scalars,  
 $y_{2t}$  and  $v_{2t}$  are  $1 \times (G_1 - 1)$  vectors,  
 $x_{1t}$  is  $1 \times K_1$ ,  
 $x_{2t}$  is  $1 \times K_2$ ,  
 $\pi_{11}$  is  $K_1 \times 1$ ,  
 $\pi_{21}$  is  $K_1 \times (G_1 - 1)$   
 $\pi_{12}$  is  $K_2 \times 1$ ,  
 $\pi_{22}$  is  $K_2 \times (G_1 - 1)$ .

Let

$$X = (X_1, X_2) = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \vdots \\ x_{1n} & x_{2n} \end{bmatrix}.$$

Then we know there exists a  $K \times K$  matrix  $Q$  such that

$$Q'X'XQ = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & nI_{K_2 \times K_2} \end{bmatrix}.$$



Introducing the transformation

$$x_t = (x_{1t}, x_{2t}) = (x_{1t}, z_{2t}) Q^{-1} = z_t Q^{-1}, \quad t = 1, \dots, n, \quad (7.3)$$

in 7.2 we obtain

$$\begin{aligned} y'_t &= \begin{bmatrix} \pi'_{11} & \pi'_{12} \\ \pi'_{21} & \pi'_{22} \end{bmatrix} x'_t + v'_t \\ &= \begin{bmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{bmatrix}' Q^{-1'} z'_t + v'_t \\ &= \begin{bmatrix} P_{11} & P_{21} \\ P_{12} & P_{22} \end{bmatrix}' z'_t + v'_t. \end{aligned} \quad (7.4)$$

The effect of the above transformation on the original model is to reparametrize the reduced form. Now, by the original specification

$$\begin{aligned} P_{11} - P_{21} \beta &= \gamma, \\ P_{12} - P_{22} \beta &= 0, \\ v_1 - V_2 \beta &= u_1, \end{aligned} \quad (7.5)$$

where  $v_1$  and  $V_2$  are a  $n \times 1$  vector and a  $n \times (G_1 - 1)$  matrix of reduced form errors. If we estimate the transformed reduced form parameters by simple least squares we obtain

$$\begin{bmatrix} \hat{P}_{11} & \hat{P}_{21} \\ \hat{P}_{12} & \hat{P}_{22} \end{bmatrix} = \begin{bmatrix} (x_1' x_1)^{-1} x_1' (y_1, y_2) \\ \frac{1}{n} z_2' (y_1, y_2) \end{bmatrix}. \quad (7.6)$$

Placing the estimates of  $P_{12}$  and  $P_{22}$  in the second equation of 7.5 and applying again, simple least squares, we obtain

$$\hat{\beta} = (\hat{P}_{22}' \hat{P}_{22})^{-1} (\hat{P}_{22}' \hat{P}_{12}) \quad (7.7)$$

We shall show that 7.7 is the two stage least squares estimator of  $\beta$ .

The two stage least squares estimator is given by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \hat{Y}_2' \hat{Y}_2 & Y_2' X_1 \\ X_1' Y_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2' y_1 \\ X_1' y_1 \end{bmatrix}.$$

Let

$$A = \hat{Y}_2' \hat{Y}_2 - Y_2' Z (Z' Z)^{-1} Z' Y_2 = \frac{1}{n} Y_2' Z_2' Z_2 Y_2, \text{ and}$$

$$B = Y_2' X_1 (X_1' X_1)^{-1}.$$

Then

$$\begin{aligned} \hat{\beta} &= A^{-1} \hat{Y}_2' y_1 - A^{-1} B X_1' y_1 \\ &= A^{-1} \{ Y_2' Z (Z' Z)^{-1} Z' y_1 - Y_2' X_1 (X_1' X_1)^{-1} X_1' y_1 \} \\ &= A^{-1} \{ Y_2' X_1 (X_1' X_1)^{-1} X_1' y_1 + \frac{1}{n} Y_2' Z_2 Z_2' y_1 - Y_2' X_1 (X_1' X_1)^{-1} X_1' y_1 \} \\ &= A^{-1} \{ \frac{1}{n} Y_2' Z_2 Z_2' y_1 \} \\ \hat{\beta} &= \{ \frac{1}{n} Y_2' Z_2 Z_2' Y_2 \}^{-1} \{ \frac{1}{n} Y_2' Z_2 Z_2' y_1 \}. \end{aligned} \quad (7.8)$$

Using the result in 7.6, expression 7.7 follows immediately.

Let us index the y's as follows:

$$Y = (y_1, y_2) = (y_1, y_2, \dots, y_{G_1})$$

We may also express

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{j1} & \cdots & p_{G_1 1} \\ p_{12} & p_{22} & \cdots & p_{j2} & \cdots & p_{G_1 2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{1K_1} & p_{2K_1} & \cdots & p_{jK_1} & \cdots & p_{G_1 K_1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{1K} & p_{2K} & \cdots & p_{jK} & \cdots & p_{G_1 K} \end{bmatrix} \quad .$$

The system of equations  $P_{12} - P_{22}\beta = 0$  can be written

$$p_{1j} = \sum_{s=2}^{G_1} \beta_s p_{sj}, \quad j = K_1 + 1, K_1 + 2, \dots, K. \quad (7.9)$$

When the reduced form coefficients are estimated by least squares we have the estimated vectors

$$\hat{P}_{\cdot j} = (\hat{p}_{1j}, \hat{p}_{2j}, \dots, \hat{p}_{G_1 j}) , \quad j = K_1 + 1, K_1 + 2, \dots, K. \quad (7.10)$$

Let us assume that the  $n \times K$  matrix of exogenous variables is a matrix of constants and is of full rank. Under this hypothesis, the vector

$$\hat{P}_{\cdot j} - P_{\cdot j} = (\hat{p}_{1j} - p_{1j}, \hat{p}_{2j} - p_{2j}, \dots, \hat{p}_{G_1 j} - p_{G_1 j}) , \quad j = K_1 + 1, \dots, K. \quad (7.11)$$

has a multivariate normal distribution with zero mean and covariance matrix  $\frac{1}{n} \phi$  where  $\phi$  is the covariance matrix of reduced form errors. That is

$$\phi = E(v'_t v_t) = \begin{bmatrix} \phi_{(11)} & \phi_{(21)} \\ \phi_{(12)} & \phi_{(22)} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{11} & \cdots & \phi_{1G_1} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2G_1} \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{G_1 1} & \phi_{G_1 2} & \cdots & \phi_{G_1 G_1} \end{bmatrix}, \quad t = 1, \dots, n \quad (7.12)$$

where  $\phi_{(11)} = \phi_{11}$ ,  $\phi_{(21)}$  is  $1 \times (G_1 - 1)$ ,  $\phi_{(12)}$  is  $(G_1 - 1) \times 1$  and  $\phi_{(22)}$

is  $(G_1-1) \times (G_1-1)$ . Thus, the problem of estimating  $\beta$  can be alternatively formulated as follows

$$\begin{aligned} p_{1j} &= \sum_{s=2}^{G_1} \beta_s p_{sj}, \quad j = K_1 + 1, \dots, K, \\ \hat{P}_{\cdot j} &= P_{\cdot j} + v_{\cdot j}, \\ v_{\cdot j} &\sim \text{NID}(0, \frac{1}{n} \phi). \end{aligned} \quad (7.13)$$

An unbiased estimator of  $\phi$  is given by

$$S = \frac{1}{n-K} \{ (y_1, y_2)' [I - X(X'X)^{-1}X'] (y_1, y_2) \}, \quad (7.14)$$

where  $S$ , distributed as a Whishart matrix with  $n - K$  degrees of freedom, is independent of  $\hat{P}$ . Equations 7.13 and 7.14 are seen to define an errors-in-variables problem completely analogous to that given in Section C of Chapter IV. The covariance matrix of the errors  $v_{\cdot j}$ ,  $j = K_1 + 1, \dots, K$  is  $\frac{1}{n} \phi$ , where  $\phi$  is given by 7.12. In addition, the elements of the second equation 7.13 are of  $O_p(1/n)$ .

If in correspondence to the partition  $(y_1, y_2)$  we write

$$S = \begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix}.$$

It can be shown that the maximum likelihood estimation of  $\beta$  is given by

$$\tilde{\beta} = [\hat{P}'_{22} \hat{P}_{22} - \lambda_1 S_{22}]^{-1} [\hat{P}'_{22} \hat{P}_{12} - \lambda_1 S_{12}] \quad (7.15)$$

where  $\lambda_1$  is the smallest root of

$$|\hat{P}'\hat{P} - \lambda S| = 0. \quad (7.16)$$

We observe that the transforming matrix  $Q$  has been chosen so that  $\frac{1}{n} Y_2' Z_2 Z_2' Y_2$  and  $\frac{1}{n} Y_2' Z_2 Z_2' Y_1$  are  $O_p(n)$ .

In particular when there is only one  $\beta$  to be estimated, expression 7.7 reduces to

$$\hat{\beta} = \left( \sum_{j=K_1+1}^K \hat{p}_{2j}^2 \right)^{-1} \left( \sum_{j=K_1+1}^K \hat{p}_{1j} \hat{p}_{2j} \right) \quad (7.17)$$

### B. The Experimental Procedure

The basic model to be considered is given by the system

$$\begin{aligned} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} &= u_{1t} \\ y_{1t} + \beta_{22} y_{2t} &+ \gamma_{23} x_{3t} + \gamma_{24} x_{4t} = u_{2t} \end{aligned} \quad (7.18)$$

previously used by Summers (1965). The reduced form of the model

$$\begin{aligned} y_{1t} &= \pi_{11} x_{1t} + \pi_{12} x_{2t} + \pi_{13} x_{3t} + \pi_{14} x_{4t} + v_{1t} \\ y_{2t} &= \pi_{21} x_{1t} + \pi_{22} x_{2t} + \pi_{23} x_{3t} + \pi_{24} x_{4t} + v_{2t} \end{aligned} \quad (7.19)$$

Assume we have a sample of size  $n$  of real data; i.e., a  $n \times 4$  matrix whose  $t$ th row is  $(x_{1t}, x_{2t}, x_{3t}, x_{4t})$ . We introduce variables

$$z_{3t} = \frac{x_{3t} - b_1 x_{1t} - b_2 x_{2t}}{\sqrt{(R.S.S.)_{3.12}}} \sqrt{n} \quad ,$$

and

$$z_{4t} = \frac{x_{4t} - c_1 x_{1t} - c_2 x_{2t} - c_3 x_{3t}}{\sqrt{(R.S.S.)_{4.123}}} \sqrt{n} \quad ;$$

where

$(R.S.S.)_{3.12}$  = Residual sum of squares obtained by regressing  $x_{3t}$  on  $x_{1t}$  and  $x_{2t}$  ;

$(R.S.S.)_{4.123}$  = Residual sum of squares obtained by regressing  $x_{4t}$  on  $x_{1t}$  ,  $x_{2t}$  and  $x_{3t}$  ;

$b_1$  and  $b_2$  = Regression coefficients estimated by regressing  $x_{3t}$  on  $x_{1t}$  and  $x_{2t}$  ;

$c_1$ ,  $c_2$  and  $c_3$  = Regression coefficients estimated by regressing  $x_{4t}$  on  $x_{1t}$  ,  $x_{2t}$  and  $x_{3t}$  .

Thus, if we want to estimate  $\beta_{12}$  of the first equation of system 7.18, the set of variables  $(x_{1t}, x_{2t}, x_{3t}, x_{4t})$ ,  $t = 1, 2, \dots, n$  , will satisfy the conditions of the transformation introduced in the previous section. For the estimation of  $\beta_{22}$  the same procedure can be applied. In this case, we regress  $x_{2t}$  on  $x_{3t}$  and  $x_{4t}$  and  $x_{1t}$  on  $x_{2t}$ ,  $x_{3t}$  and  $x_{4t}$  , etc.

Let us write

$$z_{1t} = x_{1t} ,$$

$$z_{2t} = x_{2t} ,$$

$$z_{3t} = B_3 x_{3t} - B_1 x_{2t} - B_2 x_{2t} ,$$

$$z_{4t} = C_4 x_{4t} - C_1 x_{1t} - C_2 x_{2t} - C_3 x_{3t} ,$$

where

$$B_i = \sqrt{n} b_i / \sqrt{(R.S.S.)_{3.12}} , \quad i = 1, 2, 3 ; \quad b_3 = 1$$

$$C_i = \sqrt{n} c_i / \sqrt{(R.S.S.)_{4.123}} , \quad i = 1, 2, 3, 4 ; \quad c_4 = 1 .$$

Solving the above system for the  $x$ 's we have

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{B_1}{B_3} & \frac{B_2}{B_3} & \frac{1}{B_3} & 0 \\ \frac{C_1}{C_4} + \frac{B_1 C_3}{B_3 C_4}, & \frac{C_2}{C_4} + \frac{B_2 C_3}{B_3 C_4}, & \frac{C_3}{B_3 C_4}, & \frac{1}{C_4} \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \\ z_{4t} \end{bmatrix} \quad (7.20)$$

Placing 7.20 in 7.19 we have

$$p_{13} = \frac{1}{B_3} \pi_{13} + \frac{C_3}{B_3 C_4} \pi_{14},$$

$$p_{14} = \frac{1}{C_4} \pi_{14},$$

$$p_{23} = \frac{1}{B_3} \pi_{23} + \frac{C_3}{B_3 C_4} \pi_{24},$$

$$p_{24} = \frac{1}{C_4} \pi_{24}; \quad (7.21)$$

where

$$B_3 = \sqrt{n} / \sqrt{(R.S.S.)_{3.12}},$$

$$C_4 = \sqrt{n} / \sqrt{(R.S.S.)_{4.123}},$$

$$\frac{C_3}{B_3 C_4} = \frac{C_3}{B_3} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{(R.S.S.)_{3.12}}}. \quad (7.22)$$

Let

$$E \begin{bmatrix} v_{1t}^2 & v_{1t} v_{2t} \\ v_{2t} & v_{2t}^2 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \quad t = 1, \dots, n,$$

$$\begin{aligned}
&= \begin{bmatrix} 1 & \beta_{12} \\ 1 & \beta_{22} \end{bmatrix}^{-1} \begin{bmatrix} u_{1t}^2 & u_{1t}u_{2t} \\ u_{2t}u_{1t} & u_{2t}^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \beta_{12} & \beta_{22} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 1 & \beta_{12} \\ 1 & \beta_{22} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \beta_{12} & \beta_{22} \end{bmatrix}^{-1} .
\end{aligned}$$

Then, we can describe the sampling experiment as follows:

1. A sample of a 4 dimensional normal random variable

$$\begin{bmatrix} \hat{p}_{13} \\ \hat{p}_{14} \\ \hat{p}_{23} \\ \hat{p}_{24} \end{bmatrix} \sim N_4 \left[ \begin{bmatrix} p_{13} \\ p_{14} \\ p_{23} \\ p_{24} \end{bmatrix}, \frac{1}{n} \begin{bmatrix} \phi_{11} & 0 & \phi_{12} & 0 \\ 0 & \phi_{11} & 0 & \phi_{12} \\ \phi_{12} & 0 & \phi_{22} & 0 \\ 0 & \phi_{12} & 0 & \phi_{22} \end{bmatrix} \right]$$

was selected.

2. A sample covariance matrix was selected from a parent Wishart distribution, with degrees of freedom  $p = n - K$  and covariance matrix

$$\phi = \frac{1}{n} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} .$$

Let this sample covariance matrix be denoted by

$$V = \frac{1}{n} \begin{bmatrix} \hat{\phi}_{11} & \hat{\phi}_{12} \\ \hat{\phi}_{21} & \hat{\phi}_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} .$$



3. Define

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} \hat{\pi}_{13}^2 + \hat{\pi}_{14}^2 & \hat{\pi}_{13}\hat{\pi}_{23} + \hat{\pi}_{14}\hat{\pi}_{24} \\ \hat{\pi}_{13}\hat{\pi}_{23} + \hat{\pi}_{14}\hat{\pi}_{24} & \hat{\pi}_{23}^2 + \hat{\pi}_{24}^2 \end{bmatrix}$$

and  $\lambda_1$  to be the minimum root of

$$|A - \lambda V| = 0 \quad .$$

Then, the following estimators were calculated

$$\hat{\beta}_1 = \frac{A_{12} + L_1 V_{12}}{A_{22} + L_1 V_{22}} \quad ;$$

where

$$L_1 = \max\{0, \frac{V_{22} - A_{22}}{V_{22}}\} \quad .$$

$$\hat{\beta}_2 = \frac{A_{12} - \lambda_1 V_{12} + L_2 V_{12}}{A_{22} - \lambda_1 V_{22} + L_2 V_{22}} \quad ,$$

where

$$L_2 = \max\{0, \frac{V_{22} + \lambda_1 V_{22} - V_{12}}{V_{22}}\} \quad .$$

$$\hat{\beta}_3 = \frac{A_{12} - \lambda_1 V_{12} - A_{22}}{A_{22} - \lambda_1 V_{22} + V_{22}} \quad ,$$

$$\hat{\beta}_4 = \frac{A_{12}}{A_{22} + L_1 V_{22}} \quad .$$

$\hat{\beta}_1$  is the two stage least squares estimators modified to assure finite variance.  $\hat{\beta}_2$  is the limited information maximum likelihood estimator modified to assure finite variance.  $\hat{\beta}_3$  is limited information maximum

likelihood estimator unbiased to  $O_p(1/n)$ . And  $\hat{\beta}_4$  is a different form of the two stage least squares estimator modified to assure finite variance.

4. The above process was repeated  $N$  times and the sampling variance of the  $\hat{\beta}$ 's was estimated by

$$\text{Var}(\hat{\beta}) = \frac{1}{N} \sum_1 (\hat{\beta}_i - \bar{\hat{\beta}})^2 .$$

The values chosen for the parameters were those used by Summers (1965). A total of 16 different experiments were run. There were two classes of experiments; namely the A experiments and the B experiments, 8 of each kind. The difference between the two kinds of experiments was the level of correlation between the generating exogenous variables. Summers<sup>1</sup> provided the covariance structures of his exogenous variables for both types of experiments. In Table 1 the parameter values used in the experiments are presented. In Table 2 the z's mean squares are given. We need these values to calculate the coefficients in 7.22. In Table 3 we give the correlation matrices of the z's. In Table 4 we give the covariance matrices of the reduced form errors for every experiment.

In addition, for all the experiments

$$E \begin{bmatrix} u_{1t}^2 & u_{1t}u_{2t} \\ u_{2t}u_{1t} & u_{2t}^2 \end{bmatrix} = \begin{bmatrix} 400 & 200 \\ 200 & 400 \end{bmatrix} , \quad t = 1, \dots, n .$$

---

<sup>1</sup>Summers, R. University of Pennsylvania, Philadelphia, Pennsylvania.  
Covariance structures of the experiments. Private communication. 1970.

Here, we define an experiment to be the process of estimating any parameter,  $\beta$ ,  $N$  times in any of the equations. For example, for any of the basic experiments listed in Table 1, we could estimate  $\beta_{12}$  or  $\beta_{22}$ . A digit 1 or 2 is added to the name of the basic experiment if we are estimating  $\beta_{12}$  or  $\beta_{22}$ , respectively (we would say Experiment 1A1, 3B2, etc.).

### C. Results

The results of the investigation are summarized in Tables 5, 6, 7 and 8. Table 5 presents the experimental bias of the  $\hat{\beta}$ 's as well as the theoretical bias of the maximum likelihood estimator. Table 6 presents the sampling variance of the  $\hat{\beta}$ 's as well as the theoretical variance of the maximum likelihood estimator. In Table 7 the experimental mean square errors of the estimators are given. Table 8 is a table of counts of those experiments for which there were values if  $L_1$  and (or)  $L_2$  differing from zero.

The observed bias differed considerably from the theoretical value in two of the experiments. (Experiments 1B1 and 1B2) The bias of the two maximum likelihood estimators  $\hat{\beta}_2$  and  $\hat{\beta}_3$  also differed from the theoretical value in Experiments 1A1 and 4B1, respectively. The evidence is not sufficient to declare the modified two stage least squares estimator less biased than the maximum likelihood estimator.

The modification to produce finite variance was not required in the A experiments. Of the 200 samples in Experiments 1B1, 2B1, 3B1 and 4B1. 11, 1, 10 and 7 samples, respectively, had either  $L_1$  or  $L_2$  or both  $L_1$  and  $L_2$  not equal to zero.

$\hat{\beta}_3$  showed, on the average, the smallest sampling variances and  $\hat{\beta}_2$  the largest variances. It is also evident from Table 6 that, on the average, the variances were larger for the B experiments. The experimental variances of  $\hat{\beta}_1$  were the closest to those of  $\hat{\beta}_3$ . Similar conclusions can be obtained from inspection of the mean squares in Table 7. The last column of Table 7, indicates that on the average  $\hat{\beta}_1$  was less efficient than  $\hat{\beta}_3$ . The efficiency of  $\hat{\beta}_1$  relative to  $\hat{\beta}_3$  was even smaller for the B experiments. From Table 7 we obtain the following figures:

Relative Efficiency of $\hat{\beta}_1$ to $\hat{\beta}_3$	
Experiments A	0.985
Experiments B	0.922
Average	0.954

Therefore, on the basis of the results of this sampling experiment,  $\hat{\beta}_3$  is to be preferred to the other estimators.

Table 1. Parameter combinations used in the sampling experiments

Experiment	First equation			Second equation			n	N
	$\beta_{12}$	$\gamma_{11}$	$\gamma_{12}$	$\beta_{22}$	$\gamma_{23}$	$\gamma_{24}$		
1A, 1B	-0.7	0.8	0.7	0.4	0.6	-0.4	20	200
2A, 2B	-0.7	0.8	0.7	0.4	0.6	-0.4	40	200
3A, 3B	-0.1	0.8	0.7	0.4	0.6	-0.4	20	200
3A, 4B	-1.3	0.8	0.7	0.4	0.6	-0.4	20	200

Table 2. Mean squares of the z variables

		$z_1$	$z_2$	$z_3$	$z_4$
Experiment	$z_1$	1017.94	121.99	38.40	1861.42
	$z_2$	121.99	2402.91	6.27	-428.99
	$z_3$	38.40	6.27	56.57	357.99
	$z_4$	1861.42	-428.99	357.99	23572.12
Experiment	$z_1$	1017.94	-576.69	205.63	4786.69
	$z_2$	-576.93	2402.91	-191.72	-3267.74
	$z_3$	205.63	-191.72	56.57	957.31
	$z_4$	4786.69	-3267.74	957.31	23572.12

Table 3. Intercorrelations of the z variables

		$z_1$	$z_2$	$z_3$	$z_4$
A Experiment	$z_1$	1	0.078	0.16	0.38
	$z_2$	0.078	1	0.017	-0.057
	$z_3$	0.16	0.017	1	0.31
	$z_4$	0.38	-0.057	0.31	1
B Experiment	$z_1$	1	-0.37	0.86	0.98
	$z_2$	-0.37	1	-0.52	-0.43
	$z_3$	0.86	-0.52	1	0.83
	$z_4$	0.98	-0.43	0.83	1

Table 4. Covariance matrices of reduced form residuals

Experiment	$\phi_{11}$	$\phi_{12}$	$\phi_{22}$
1A and 1B	307.4	49.6	330.6
2A and 2B	307.4	49.6	330.6
3A and 3B	336.0	-240.0	1600.0
4A and 4B	328.0	62.3	138.4

Table 5. Experimental and theoretical bias of the  $\hat{\beta}$ 's

Experiment	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Theoretical M.L.E.	$1.96 \hat{\sigma}^1$
1A1	+0.00408	+0.01041*	+0.00443	+0.00408	-0.00581	0.01492
2A1	-0.00132	+0.00179	-0.00115	-0.00132	-0.00291	0.01054
3A1	+0.00462	+0.00430	+0.00467	+0.00462	+0.00034	0.00490
4A1	+0.00958	-0.00075	+0.01154	+0.00959	-0.01196	0.02680
1B1	+0.07991	-0.03830*	+0.11191	+0.07829	-0.10854	0.06450
2B1	+0.00751	-0.07849	+0.00645	+0.00751	-0.05427	0.04560
3B1	+0.00598	+0.00468	-0.00020	+0.00660	+0.00638	0.02202
4B1	+0.09250	-0.16105	+0.14587	+0.09274	-0.22341	0.11560
1A2	+0.00119	+0.00376	+0.00108	+0.00119	+0.00277	0.01466
2A2	-0.00010	+0.00110	-0.00023	-0.00010	+0.00138	0.01036
3A2	+0.00244	+0.00242	+0.00014	+0.00756	+0.00609	0.01044
4A2	+0.01447	+0.01328	+0.01349	+0.01447	-0.00056	0.02190
1B2	+0.01676	+0.02350	+0.01648	+0.01676	+0.00514	0.01996
2B2	-0.00984	-0.00741	-0.01010	-0.00984	+0.00257	0.01408
3B2	-0.01256	+0.00068	-0.01134	-0.01256	+0.01132	0.01422
4B2	-0.03257	-0.03453*	-0.03259	-0.03257	+0.00103	0.02980

$^1(\hat{\sigma})^2$  = Theoretical variance of the maximum likelihood estimator  
divided by N.

\* Estimated bias differs from the theoretical by more than  $1.96 \hat{\sigma}$ .

Table 6. Experimental and theoretical variances of the  $\hat{\beta}$ 's

Experimental	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Theoretical M.L.E.
1A1	0.01031	0.01026	0.00994	0.01031	0.01114
2A1	0.00619	0.00637	0.00629	0.00619	0.00557
3A1	0.00136	0.00138	0.00135	0.00136	0.00119
4A1	0.03194	0.03237	0.03140	0.03194	0.03575
1B1	0.18827	0.28266	0.15171	0.18830	0.20811
2B1	0.12556	0.17705	0.09832	0.12556	0.10405
3B1	0.03248	0.07404	0.03539	0.03196	0.02426
4B1	0.72588	1.13272	0.51576	0.73042	0.66774
1A2	0.01488	0.01514	0.01476	0.01488	0.01075
2A2	0.00544	0.00547	0.00539	0.00544	0.00537
3A2	0.00516	0.00523	0.00498	0.00516	0.00543
4A2	0.02498	0.02519	0.02474	0.02498	0.02404
1B2	0.02005	0.02130	0.01987	0.02005	0.01996
2B2	0.01105	0.01122	0.01092	0.01105	0.00998
3B2	0.00954	0.01047	0.00943	0.00954	0.01013
4B2	0.05100	0.05258	0.04989	0.05100	0.04465



Table 7. Experimental mean square errors of the  $\hat{\beta}$ 's

Experiment	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	MSE $\hat{\beta}_3$ /MSE $\hat{\beta}_1$
1A1	0.01033	0.01037	0.00996	0.01033	0.96
2A1	0.00619	0.00637	0.00629	0.00619	1.01
3A1	0.00138	0.00140	0.00137	0.00138	0.99
4A1	0.03203	0.03237	0.03153	0.03203	0.98
1B1	0.19466	0.28413	0.16423	0.19443	0.84
2B1	0.12562	0.17711	0.09836	0.12562	0.78
3B1	0.03252	0.07406	0.03539	0.03200	1.09
4B1	0.73444	1.15462	0.53704	0.73903	0.73
1A2	0.01488	0.01515	0.01476	0.01488	1.00
2A2	0.00544	0.00547	0.00539	0.00544	0.99
3A2	0.00517	0.00524	0.00498	0.00522	0.96
4A2	0.2519	0.02537	0.02492	0.02519	0.99
1B2	0.02033	0.02185	0.02014	0.02033	0.99
2B2	0.01115	0.01127	0.01102	0.01115	0.98
3B2	0.00970	0.01047	0.00956	0.00970	0.99
4B2	0.05206	0.05377	0.05095	0.05206	0.98

Table 8. Table of counts of  $\hat{\beta}_1$  and  $\hat{\beta}_3$  conditional on values of  $L_1$  and  $L_2$ 

	$L_1 = 0$ $L_2 = 0$	$L_1 = 0$ $L_2 \neq 0$	$L_1 \neq 0$ $L_2 = 0$	$L_1 \neq 0$ $L_2 \neq 0$	Total
Experiment 1B1					
$ \hat{\beta}_1 - \beta  >  \hat{\beta}_3 - \beta $	86	3	0	2	91
$ \hat{\beta}_3 - \beta  >  \hat{\beta}_1 - \beta $	103	3	0	3	109
Total	189	6	0	5	200
Experiment 2B1					
$ \hat{\beta}_1 - \beta  >  \hat{\beta}_3 - \beta $	100	1	0	0	101
$ \hat{\beta}_3 - \beta  >  \hat{\beta}_1 - \beta $	99	0	0	0	99
Total	199	1	0	0	200
Experiment 3B1					
$ \hat{\beta}_1 - \beta  >  \hat{\beta}_3 - \beta $	108	2	0	6	116
$ \hat{\beta}_3 - \beta  >  \hat{\beta}_1 - \beta $	82	2	0	0	84
Total	190	4	0	6	200
Experiment 4B1					
$ \hat{\beta}_1 - \beta  >  \hat{\beta}_3 - \beta $	89	3	0	2	94
$ \hat{\beta}_3 - \beta  >  \hat{\beta}_1 - \beta $	104	1	0	1	106
Total	193	4	0	3	200

## VIII. SUMMARY

Alternative estimators for the errors in variables problem are investigated. Fuller's (1968) estimator,  $\tilde{\beta}$ , a modification of the moment estimator,  $\hat{\beta}$ , that is unbiased to  $O_p(1/n)$  is investigated. The general conditions under which to  $O_p(1/n^2)$   $\tilde{\beta}$  has smaller mean square error than  $\hat{\beta}$  are established. In the important case when  $\beta$  is a scalar, it is shown that  $\tilde{\beta}$  has a smaller mean square error than  $\hat{\beta}$ . The expected values of the estimators  $\hat{\beta}$  and  $\tilde{\beta}$  as constructed in Chapter IV, are not necessarily defined for finite samples. A method to achieve finite moments is suggested. An application of the techniques to the estimation of a single equation in a system of equations is illustrated. The estimation of some functions of the parameters of a regression model, found frequently in economic studies, is also illustrated.

Contributions to the knowledge of the asymptotic properties of the maximum likelihood estimator are obtained. These results together with those of Fuller (1970b) provide a better understanding of the errors in variables problem. An immediate consequence of these results is the derivation of an estimator unbiased to  $O_p(1/n)$  with smaller mean square error than the moment estimator.

A review and extension of DeGracie's (1968) work on a covariance model with errors in the covariates is presented. Using generalized inverses the methodology is extended to a nonfull rank covariance model. It is shown that the system of normal equations can be solved in a way completely analogous to the standard covariance model. It is observed that the

maximum likelihood estimation procedure can also be applied.

A sampling experiment was used to compare the properties of several estimators. The bias and relative efficiency of the estimators were a function of the parametric configuration. The modified maximum likelihood estimation unbiased to  $O_p(1/n)$  was uniformly superior to the maximum likelihood estimator modified only to have finite variance. For the eight parametric configurations investigated the relative efficiency of the two stage least squares estimator (modified to have finite variance) to the maximum likelihood estimator (modified to be unbiased to  $O_p(1/n)$ ) was 0.954.

## IX. BIBLIOGRAPHY

- Adcock, R. J. 1877. A problem in least squares. *Analyst* 5: 53-54.
- Arnold, B. 1968. Mimeographed class notes, Statistics 642, Winter 1967-68. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.
- Bodewig, E. 1956. *Matrix calculus*. Amsterdam, Holland, North-Holland Publishing Company.
- Chernoff, H. 1956. Large sample theory: parametric case. *Mathematical Statistics Annals* 27: 1-22.
- Chung, K. L. 1968. *A course in probability theory*. New York, New York, Harcourt, Brace and World, Incorporated.
- DeGracie, J. S. 1968. Analysis of covariance when the concomitant variable is measured with error. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Dorff, M. 1960. Large and small sample properties of estimators for a linear functional relationship. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Fuller, W. A. 1968. The theory of econometrics. Unpublished class notes. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.
- Fuller, W. A. 1970a. Mimeographed class notes, Statistics 638, Winter 1969-70. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.
- Fuller, W. A. 1970b. Unpublished research. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.
- Goldberger, A. 1963. *Econometric theory*. New York, New York, John Wiley and Sons, Incorporated.
- James, A. T. 1954. Normal multivariate analysis and the orthogonal group. *Mathematical Statistics Annals* 25: 40-75.
- Johnston, J. 1963. *Econometric methods*. New York, New York, McGraw-Hill Book Company, Incorporated.
- Kendall, M. G. 1951. Regression, structure and functional relationship. Part 1. *Biometrika* 38: 11-25.

- Kendall, M. G. 1952. Regression, structure and functional relationship. Part 2. *Biometrika* 39: 96-108.
- Kummell, C. H. 1879. Reduction of observation equations which contain more than one observed quantity. *Analyst* 6: 97-105.
- Lindley, D. V. 1947. Regression lines and the linear functional relationship. *Royal Statistical Society Journal* 9, Supplement: 219-244.
- Madansky, A. 1959. The fitting of straight lines when both variables are subject to error. *American Statistical Association Journal* 54: 173-205.
- Malinvaud, E. 1966. Statistical methods of econometrics. Chicago, Illinois, Rand McNally and Company.
- Mann, H. B. and Wald, A. 1943. On stochastic limit and order relationships. *Mathematical Statistics Annals* 14: 217-226.
- Meier, P. 1953. Variance of a weighted mean. *Biometrics* 9: 59-73.
- Nagar, A. L. 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27: 575-595.
- Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2: 559-572.
- Pratt, J. W. 1959. On a general concept of "in probability". *Mathematical Statistics Annals* 30: 549-558.
- Reiersol, O. 1945. Confluence analysis by means of instrumental sets of variables. *Arkiv for Matematik, Astronomi och Fysik* 32: 1-119.
- Reiersol, O. 1950. Identifiability of a linear relation between variables which are subject to error. *Econometrica* 18: 375-389.
- Summers, R. 1965. A capital intensive approach to the small sample properties of various simultaneous equation estimators. *Econometrica* 33: 1-41.
- Tucker, H. G. 1967. A graduate course in probability. New York, New York, Academic Press, Incorporated.
- Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11: 284-300.
- Widder, D. V. 1961. Advanced calculus. Englewood Cliffs, New Jersey, Prentice-Hall, Incorporated.

Wolfowitz, J. 1953. Estimation by the minimum distance method. Institute of Statistical Mathematics Annals 5: 9-23.

Zyskind, G. 1964. Mimeographed class notes, Statistics 641, Fall 1964. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.

## X. ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Wayne Fuller for his helpful suggestions and critical remarks during the course of this investigation.

Special thanks are given to Eric West who prepared the program for the sampling experiment of Chapter VII and to Mrs. Shirley Saveraid for her excellent typing of the manuscript. Thanks are also given to Dr. T. A. Bancroft for his interest in the author's work while staying at Iowa State University.

The author is indebted to the Ford Foundation for supporting his graduate work, to the Colegio de Postgraduados from Chapingo, Mexico, in particular to the Centro de Estadística y Cálculo for much help and to Iowa State University for the provision of funds for computer use.

Finally, the author acknowledges the influence of his wife, Maria Elena, and children as a source of inspiration in his work.